NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

CROWDSOURCED FORMAL VERIFICATION: A BUSINESS CASE ANALYSIS TOWARD A HUMAN-CENTERED BUSINESS MODEL

by

Andreas Baur

June 2015

Thesis Advisor:	Geoffrey G. Xie
Second Reader:	Nicholas Dew

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	06-19-2015	Master's Thesis 12-12-2013 to 05-14-2015

4. TITLE AND SUBTITLE

CROWDSOURCED FORMAL VERIFICATION: A BUSINESS CASE ANALYSIS TOWARD A HUMAN-CENTERED BUSINESS MODEL

5. FUNDING NUMBERS

6. AUTHOR(S)

Andreas Baur

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Postgraduate School
Monterey, CA 93943

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

N/A

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views expressed in this document are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol Number: N/A.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

The DARPA project Crowd Sourced Formal Verification (CSFV) tries to investigate whether offering free games via the Internet that translate player's actions into program annotations helps to overcome the challenges of the expensive and time-consuming formal verification of software by human experts. This business case analysis evaluates the results of the CSFV-project phase 1. Based on data of the games, the author identifies three problems of the current CSFV approach. The author concludes, in accordance with the Gartner Hype Cycle Research Methodology, that the technology currently is not sufficiently mature to justify a financial investment, but that the cutting-edge approach may reach the plateau of productivity within two to five years, due to parallel maturation of some technologies. The author argues that a human-centered approach is necessary to transform the customer base in order to mitigate the identified deficiencies and to leverage crowdsourced formal verification as a sustainable business. He first explains the concepts relevant in the context of crowdsourced formal verification and the technologies having impact on it. He then identifies the current issues and existing obstacles in the current technology. Based on future trends and visions in the respective fields of technology, and the needs and motivations of people, he proposes a human-centered business model that may foster the implementation of crowdsourced formal verification of software in organizations that depend on security-critical and safety-critical software.

14. SUBJECT TERMS	15. NUMBER OF PAGES
human behavior, scripting, dynamic behavior, knowledge representation, ontology, Protégé, COMBATXXI	105
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

CROWDSOURCED FORMAL VERIFICATION: A BUSINESS CASE ANALYSIS TOWARD A HUMAN-CENTERED BUSINESS MODEL

Andreas Baur
Commander, German Navy
Diplom-Kaufmann (univ.), Universität der Bundeswehr München, 2000

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF BUSINESS ADMINISTRATION

from the

NAVAL POSTGRADUATE SCHOOL
June 2015

Author: Andreas Baur

Approved by: Geoffrey G. Xie
Thesis Advisor

Nicholas Dew
Second Reader

William Gates
Chair, Graduate School of Business and Public Policy

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The DARPA project Crowd Sourced Formal Verification (CSFV) tries to investigate whether offering free games via the Internet that translate player's actions into program annotations helps to overcome the challenges of the expensive and time-consuming formal verification of software by human experts. This business case analysis evaluates the results of the CSFV-project phase 1. Based on data of the games, the author identifies three problems of the current CSFV approach. The author concludes, in accordance with the Gartner Hype Cycle Research Methodology, that the technology currently is not sufficiently mature to justify a financial investment, but that the cutting-edge approach may reach the plateau of productivity within two to five years, due to parallel maturation of some technologies. The author argues that a human-centered approach is necessary to transform the customer base in order to mitigate the identified deficiencies and to leverage crowdsourced formal verification as a sustainable business. He first explains the concepts relevant in the context of crowdsourced formal verification and the technologies having impact on it. He then identifies the current issues and existing obstacles in the current technology. Based on future trends and visions in the respective fields of technology, and the needs and motivations of people, he proposes a human-centered business model that may foster the implementation of crowdsourced formal verification of software in organizations that depend on security-critical and safety-critical software.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

CDF	Cumulative Distribution Function
CSFV	Crowd Sourced Formal Verification
CT	computerized axial tomography
DARPA	Defense Advanced Research Projects Agency
DAU	Daily Average User
EPC	Electronic Product Code
ER	Engagement Rate
DOD	Department of Defense
FRAMA-C	Framework for Modular Analysis of C programs
IOT	Internet of Things
KLOC	Thousands (Kilos) of Lines of Code
MAU	Monthly average user
MILEs	Massive Interactive Learning Events
MS	Microsoft
NASA	National Aeronautics and Space Administration
NPS	Naval Postgraduate School
OECD	Organisation for Economic Co-operation and Development
PPSR	Public Participation in Scientific Research
RFID	Radio-frequency identification
RTE	Runtime Error

SMT	Satisfiability Modulo Theories
US	United States
USG	United States government

Acknowledgments

I would like to express my sincere gratitude to my advisor, Professor Geoffrey G. Xie from the Computer Science department, for his unlimited support, patience, and encouragement throughout the last eight months. It is not often that one finds an advisor who encourages one to tread this unknown path from business administration to computer science.

I would also like to express my sincere gratitude to Professor Nicholas Dew from the business school who taught me not to restrict my thinking to the small picture, who always inspired me to look further for something new, and who pushed my hunger for knowledge to extend.

I thank Charles Prince and Justin Rohrer for listening to my ideas, and for the long discussions that followed. My thanks go also to Umit Tellioglu, who aroused my curiosity about the Crowdsourced Formal Verification (CSFV), and Mehmet Yilmaz, who accompanied me throughout the thesis project.

I thank the Defense Advanced Research Projects Agency and the CSFV team around program manager Dr. Mike Hsieh, and his predecessor Dr. Daniel Ragsdale, who supported my research by all means.

I am fortunate to have met the great team of Galois Inc., Portland. Having had the opportunity to learn from the experts in the field and to study the relevant theoretical frameworks and paradigms has been worth its weight in gold. My special thanks goes to Rob Wiltbank who allowed me to spend time at their headquarters, and who taught me to focus on sellable products. I also would like to thank Jef Bell for giving me the opportunity to meet Aaron Tomb and Simon Winwood. Aaron and Simon set the foundation of my understanding in the field of software correctness. Their guidance, patient discussions, provision of data, and endless answers allowed me to come so far. A final thanks goes to Dylan McNamee, who even interrupted his vacation to share his insights with me.

Nothing would have been possible without my wife Rebekka's unwavering love and support, who encouraged me even in the hardest moments of my research. Without her endless love and support, this thesis would not have been completed.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction, Research Objectives and Thesis Organization

> A "bug," which in computing can occur in hardware or software, causes unexpected or unintended behavior that diverges from the product's specification. [...] While bugs can be avoided to a certain extent by carefully planning and designing using the established software development processes and by practicing good code style, the more complex a program gets, the more likely it is to contain bugs.
>
> —David Padua, Encyclopedia of parallel computing, 2011

As early as 1999, the U.S. president's IT Advisory Committee stated that "we have become dangerously dependent on large software systems whose behavior is not well understood and which often fail in unpredictable ways" (Gray, 1999). Examples illustrate the influence of errors and their deadly consequences, like the 1996 Ariane 5 maiden flight explosion,[1] the 2009 computerized axial tomography (CT) brain scan overdose in Los Angeles,[2] and the defective Toyota Camry ETCS-i system in 2005.[3]. The influence on mission-critical operations is demonstrated by examples like the 2013 National Aeronautics and Space Administration (NASA) Curiosity Mars rover standby[4] and the 911 phone service outage

[1]The Ariane 5 satellite launcher malfunction was caused by a faulty software exception routine resulting from a bad 64-bit floating point to 16-bit integer conversion.

[2]Two hundred and six patients who underwent CT scans at Cedars-Sinai Medical Center in Los Angeles were exposed to radiation overdoses. Hospital officials stated that a computer-resetting error caused the overdoses (Chitale, 2009).

[3]The ETCS-i (Electronic Throttle Control System-intelligent) is a system that uses a computer to electrically control the throttle valve opening. In a lawsuit claiming a defect in a Camry caused the vehicle to unintentionally accelerate, leading to an accident that left one woman dead and another injured, an Oklahoma jury found the Camry's electronic system was defective (Toyota loses, 2013). However, Exponent, a company ordered to analyze the system, concluded that based on their investigation, the electronics and software were not the root cause of the reported incidents of unintended acceleration in the evaluated Toyota vehicles (Exponent Inc., 2012).

[4]In this case, the NASA experts concluded a hardware failure due to its non-volatile memory, probably related to the hardware's age. The standby was fixed by changing the software code (BBC News, 2014).

on April 9, 2014.[5]

With an increasing amount of mission critical software dependency, the need for software verification becomes ever more important. Formal verification became popular in making sure that software functions as specified without producing unexpected results. Traditionally, only human experts do formal verification. Formal software verification is an arduous, time consuming, and complicated process that requires a wide variety of skills. The experts use deductive methods, like model checking, or static analysis to formally verify the correctness of the software code properties.

While formal verification is very effective (e.g., one can achieve to have only between 0.1 and 0.5 bugs per Thousands (Kilos) of Lines of Code (KLOC)) it is also extremely expensive. Software development costs increase by 2x to 100x (e.g., the seL4 microkernel formal verification took 11 person-years) (Dean, 2011). Additionally, there are only approximately 1,000 human experts in the U.S. available, and about 4,000 experts worldwide (Dean, 2011). Fundamental formal verification problems still resist automation. While heuristics have improved, they are still incomplete and do not allow full automation. As human software verification experts are a scarce resource, and budgetary constraints shape the employment of high paid resources, new ways of effective verification are examined by the Defense Advanced Research Projects Agency (DARPA).

1.1 The DARPA Crowdsourced Formal Verification Project

One potential direction to mitigate the dependency of formal software verification on human experts and budgetary constraints is crowdsourced formal verification. The DARPA explores whether and how non-experts can contribute to software verification by playing free games on the Internet that are specifically developed for this purpose.[6]

Different as the human-expert-only verification process, the crowdsourced formal verification approach is, defined through an alternating interaction between humans and machines. In general, game builders automatically convert code fragments that need to be verified

[5]Nearly 11 million people in seven states lost access to emergency services when a software programming error resulted a six-hour long 911 outage (Kieler, 2014; Federal Communications Commission, 2014).

[6]Von Ahn (2006) coined the term "Games with a purpose." He argued that games are "a seductive method for encouraging people to participate in" collective computation. Tellioglu, Xie, Rohrer, and Prince (2014) classified such crowd-sourcing efforts into a new genre called Crowd-Sourced Serious Games.

against specified security flaws into different visual problems or game instances. If players solve these problems, they effectively produce mathematical equations. These equations are called assertions, which then will be used to check whether the specifications of the games hold true.

The project Crowd Sourced Formal Verification (CSFV) aims to investigate whether large numbers of non-experts can perform formal verification faster and more cost-effectively than traditional formal verification done by human experts (DARPA, 2013). DARPA also wants to find out, which solutions non-experts would find that a computer cannot find so far. In the rest of this thesis, I use the abbreviation CSFV for the DARPA project and distinguish it from crowdsourced formal software verification as a general concept.

1.2 The Verification Games

Under the DARPA CSFV program, five teams from academia, private sector research, and program developing business found different approaches to transform formal verification into games that embrace users to solve the difficult verification problems for fun. The games are offered free-to-play on the Internet platforms Verigames.com and Verigames.org providing a crowd-sourced contribution to the verification of C-language programs. I further refer to these games as the Verigames. During the first phase from December 2013 to September 2014, Verigames.com offered four online browser games, Circuitbot, Stormbound, Ghostmap, FlowJam, and one iOS game, Xylem—the code of plants (Verigames.com, 2015). In the second phase, which started in May 2015, the teams added five more games, which were developed on basis of the lessons identified in the first phase. Players need to be full-aged due to government regulations regarding volunteer participants, but do not need to register. By playing the games, open source programs are being used by the Defense Department and other governmental and commercial organizations are reviewed. In this thesis, we focus on the games of the first phase, which I briefly introduce at this point. A more detailed analysis of some of the games will be presented in Chapter 3.

Circuitbot is set up as a strategic resource management game. In the game, players link up a team of robots to carry out missions in order to colonize different planets. Players have to activate the links between robots in logical order to gain points (DARPA, n.d.)

Flow Jam's verification approach is based on type theory. Player have to analyze and

adjust a cable network to maximize its flow by toggling variables (in that game called widgets) individually. Players advance by finding the correct relationship between links and passages.

In Ghost Map, the player is a cybernetic entity attempting to achieve consciousness. Players are trying to find a path through a brain network. Players operate Ghost Map and move forward in the game by solving the puzzle's structure. The game uses model checking as its software verification technique. The Ghost Map project is led by Raytheon BBN Technologies with support from Breakaway Games, the University of Central Florida, and Carnegie Mellon University.

Xylem challenges the players to catalog species of plants using mathematical formulas. It is a logical induction puzzle game where the player plays a botanist exploring and discovering new forms of plant life on a mysterious island called Miraflora. The game is only for iOS on Apple iPads.

In Stormbound, players have to unweave the windstorm into patterns of streaming symbols. The puzzle game challenges players to find patterns in magical energy in order to save their planet. In the game, players educate a semi-spiritual and semi-physical entity named Gola by defining the correct relationship of two given patterns. This action charges Gola's power source and helps it to defeat the storms. The game was developed by Galois, specialists in formal methods, and voidALPHA, a video game studio. Stormbound is in the focus of our analysis in this thesis.

1.3 Purpose of this Thesis

It is still unclear how crowdsourced formal verification contributes effectively to the overall verification of code. In the best case, a large number of players contribute to the games and produce all necessary valid and useful assertions to provide assurance that the code is free of certain bugs. Previous research on the games data revealed that games with a purpose, like the CSFV games, have a lower Engagement Rate (ER) than other games (Tellioglu et al., 2014). So far, it has not been examined, whether the participating crowd may produce a higher amount of assertions than a human expert that allow a later validation and verification. It is therefore unknown how valuable the players contribution is in terms of the formal verification goals.

Assuming ideal scheduling, the DARPA project initially estimated 10 minutes for solving a game level. If true, and users play for 30 minutes every day, one property of the Guava Project may be verified in one day by only 158 users.[7] One property of Daikon may be verified in one day by only 350 users.[8] DARPA assumes that in comparison with human experts and because of latency effects (but not the level of effort) formal verification in real world would take longer (Dean, 2011). These estimates have not been verified so far, and they need to be analyzed how long it takes to produce an assertion with the games. Looking closer at produced assertions, no quantitative insights are available on how useful and beneficial the crowdsourced verification process is. The amount of useful user assertions in proportion to the amount of verification goals could explain how effectively the player contributes to the verification.

This business case analysis evaluates the results of the CSFV phase 1 games launched in December 2013. The author identifies three problems with the current CSFV approach: in the current maturity of the verification technology, the acceptance of the actual games, and the selection of the audience. However, the game developing teams decided to lever the curve for math in the games. This will further limit the potential user base and is most likely not leading towards permanent higher player numbers. Moreover, it will push the games into a niche, that may limit its attractiveness. Referring to the Gartner Hype Cycle Research Methodology, we find evidence that the technology currently is not sufficiently mature to justify[9] a financial investment, but that the cutting-edge approach may reach the plateau of productivity within two to five years, due to parallel maturation of some other technologies.

Following McLuhan's thinking on technology in general that "We become what we behold.

[7]The Guava Project, formerly known as Google Collections, is one of several of Google's core libraries that are used by Google in production services in their Java-based projects. The library contains code for collections, caching, primitives support, concurrency libraries, common annotations, string processing, I/O, and so forth (Guava Project, n.d.).

[8]The Daikon invariant detector is a product of the University of Washington and reports likely program invariants. Daikon can detect properties in C, C++, C#, Eiffel, F#, Java, Perl, and Visual Basic ("Daikon dynamic invariant detector," 2015).

[9]Maturity is thereby defined "by consumer acceptance of the basic service idea, by widespread believe that the products of most manufacturers will perform satisfactorily, and by enough familiarity and sophistication to permit" (Paliwoda & Thomas, 1998). It needs to be distinguished from technical maturity and competitive maturity.

We shape our tools and then our tools shape us" (Culkin, 1967), we argue that a human-centered approach is necessary to transform the customer base in order to mitigate the identified deficiencies and to leverage crowdsourced formal verification as a sustainable business.

We adapt the concept of Design Thinking according to Simon (1969). The author first explains the concepts relevant in the context of crowdsourced formal verification and the technologies having impact on it (Chapter 2). He then identifies the current issues and existing obstacles in the current technology (Chapter 3). Based on future trends and visions in the respective fields of technology, and the needs and motivations of today's society (Chapter 4), he proposes a human-centered business model that may foster the implementation of CSFV in organizations depending on security-critical and safety-critical software.

1.4 Research Objective

The objective of this research is

1. To identify the problem space of crowd-sourced formal verification
2. To analyze the current obstacles with the elements of crowdsourced formal verification technology.
3. To describe a potential solution to better utilize crowd-sourced formal verification in business.
4. To compare future prospects and alternatives of crowdsourcing and gamification to increase software correctness in safety- and security-critical systems.

1.5 Methodology

This research focuses on helping to decide whether the crowdsourced formal verification approach should be established based on the current state of the technology. The thesis uses a mixed-methodology to evaluate this question. Secondary sources are reviewed initially using a range of information sources such as the library and Internet databases.

Taking a systems thinking perspective, we combine a quantitative measure of abstract verification with a qualitative text-based analysis to explore the maturity of crowdsourced formal verification. The primary data collection is executed by the Stormbound back end

engineer team at Galois Inc., Portland.[10] This data includes anonymous session times, assertion counts for each program point, and the results from running the checker on the submitted assertions as recorded by the game client.

In the following data analysis phase, the collected data is examined in three steps. First, we conduct a quantitative analysis at the backend data and try to measure the usefulness of the crowdsourced formal verification approach during the phase 1 of the CSFV project. *Usefulness* is defined as "the quality of having utility and especially practical worth or applicability" (Usefulness, n.d.). In a second step, the dataset of former research (cf. Tellioglu, 2014) is being reviewed to compare the actual game impact since May 2014. Finally, based on lessons identified from the developing teams, we look at, whether the teams identified the selected niches of the games for phase 1 still appropriate for phase 2.

1.6 Thesis Organization

The reader has had a brief overview of the Crowd Sourced Formal Verification project, our approach the purpose of the thesis, and the research objective.

In Chapter 2, we dive into the concepts discussed in this thesis. We introduce the concepts of the Diffusion of Technology and the Hype Cycle Research methodology as these concepts allow a evaluation of market visibility and its relevance. The other concepts are: Formal Software Verification, Crowdsourcing, and the the gamified crowdsourced approach Crowdsourced Formal Verification examined by DAPRA.

In Chapter 3, we undertakes a problem diagnosis of the CSVF phase 1 games with focus on the Stormbound game. The author analyzes the available actual data for the formal verification process produced by the game players of the Stormbound game. It also reviews the the dataset of former research (Tellioglu, 2014) to compare the actual game impact since the release of the study.

In Chapter 4, we have a closer look on how already identified technology trends on the hype cycle curve like gamification, and crowdsourcing, and citizen science may influence or shift crowdsourced formal software verification in the future. Summarizing the insights from

[10]The term back end as used in the document describes the systems, which performs the annotation of the source code, and the verification of the assertions submitted by the game players

Chapter 3 and Chapter 4, we find that technological advances will allow implementation of more automatization to the formal software verification, but conclude that a focus on the technology only falls short in solving the games weakest point - an embracing of players. We, therefore, propose a human-centered approach that integrates an understanding for the society's mindset first.

In Chapter 5, we develop a potential solution to better utilize crowdsourced formal verification in business by applying this human-centered approach. Instead of targeting citizen scientists, we recommend a focus on the educational sector. By using the business canvas methodology, we try to give a strategic recommendation to better gain advantage of the crowdsourced formal verification approach in achieving software correctness if early adopters decide to break in that direction.

Lastly, in Chapter 6, we will draw the conclusions and provide recommendations for further studies with regards to the crowdsourced formal verification approach from a business perspective.

1.7 Limitations of the Study

The author is aware that the outcome of the thesis is preliminary. The dataset encompasses data from the release of the Stormbound game in late 2013 until today. However, two factors affect the data. First, while there was a high attraction of players, and the game has been played by over 11,809 people through the end of September 2014, and game players have solved over 23,759 problems, interest settled down to a steady-state of 62 players per week.

Second, the conclusions drawn through the Stormbound dataset may not be representative for the crowdsourced formal verification approach in general. The five games of the CSFV project phase 1, Stormbound, Flow Jam, Ghost Map, CircuitBot, and Xylem, show similarities in the used formal methods, but distinguish themselves in terms of the execution of the games. For example, Stormbound and Xylem, base their progress on finding loop as a means to formally verify the code. However, instead of challenging the player with mathematical equations, the Stormbound game presents a visual interpretation of loops through icons. Players identify and combine patterns without having any mathematical skills. Therefore, the results of the stormbound game cannot be representative for other

games per se.

Third, the research and findings are based solely on the authors personal understanding and his bias may shape the outcome and conclusions. Another evaluator may have a completely different perspective, resulting in different findings. This is especially relevant, as phase 2 of the CSFV program may add additional insights that may not have been found during phase 1 as factors like marketing efforts, publicity, and the experience of the game developing teams have continuously grown.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 2:
Theoretical Considerations on Crowdsourced Formal Verification and Technological Maturity

> [I]n most cases the more inclusive and, importantly, vague and broad church definition won the day.
>
> —Muki Haklay, Gartners hype cycle and citizen science, 2013

In the first chapter, we briefly introduced the relevance of software bugs and the implications for software, especially on safety-critical and security-critical software, and why it is important to make sure that software is relatively free of bugs. This chapter introduces the concepts used in this thesis to evaluate the market maturity and the relevance of crowdsourced formal verification games. First, we introduce different methodologies on market visibility of innovations. Market visibility has two connected effects that we assess. Market visibility attracts users and makes a product or technology worth to examining. Also, it is closely connected with the perception of how well a technology functions. In a next step, we introduce the basics of formal verification of software from a nontechnical perspective providing a necessary understanding for business people rather than computer scientists. This aims at fostering an understanding of formal software verification from a managerial perspective. It describes how to better meet business objectives by gaining advantages from formal verification to achieve code correctness, reliability, and robustness. We then switch to a concept, crowdsourcing, and describe how it already impacts formal software verification and show examples of different manifestations in academia and real world.

2.1 From Public Visibility to Expectations to Market Maturity

Market maturity of formal software verification is not a question of whether it can be done, but when it will see widespread use. To be viable for financial investment it has to reach maturity, "the stage in the product life cycle where sales growth ultimately peaks, then

slows as the product reaches widespread acceptance, and competition is fierce" (Paliwoda & Thomas, 1998). The discussion about formal verification methodology and technology never stopped, but subsided in the early 1990s due to a lack of mature automatization techniques. The issue reemerged in the early 2000s when these conditions changed and new technological approaches to automate the human effort become public.

With the introduction of the CSFV games to the market, we target a better understanding of potential market impacts of the new crowdsourced approach of formal software verification. Different methodologies have been developed to better understand how to evaluate the impact of innovations and its maturity. We introduce two of these concepts, which focus on market visibility to assess how it may attract users, and how well the new technology functions may be perceived to finally recommend a business model.

2.1.1 The Diffusion of Innovations

Rogers (1962) developed a theory on the diffusion of innovations. According to Rogers, diffusion is the process "by which an innovation is communicated through certain channels over time among the participants in a social system" (Rogers, 2010, p. 35). He argues that there is a point at which an innovation reaches critical mass within the rate of adoption. Figure 2.1 shows how Rogers distinguishes five categories of adopters :

1. Innovators (2.5%)
2. Early adopters (13.5%)
3. Early majority (34%)
4. Late majority (34%)
5. Laggards (16%)

Rogers proved that the concept of diffusion of innovation can be applied to all innovations. For him innovation is an idea, practice, or object perceived as new by the adopter. This also applies to the crowdsourced formal software verification approach, which is not a innovative technology by itself, but rather a merger of different known concepts. Adoption is an individual process, it is the critical mass, the group phenomena of diffusion that allows technology to spread (Rogers, 2010). We use this concept of innovation diffusion as it provides a better understanding of how one may expect adopters of a new technology or product, and the phases through which a new techology, like crowdsourced formal

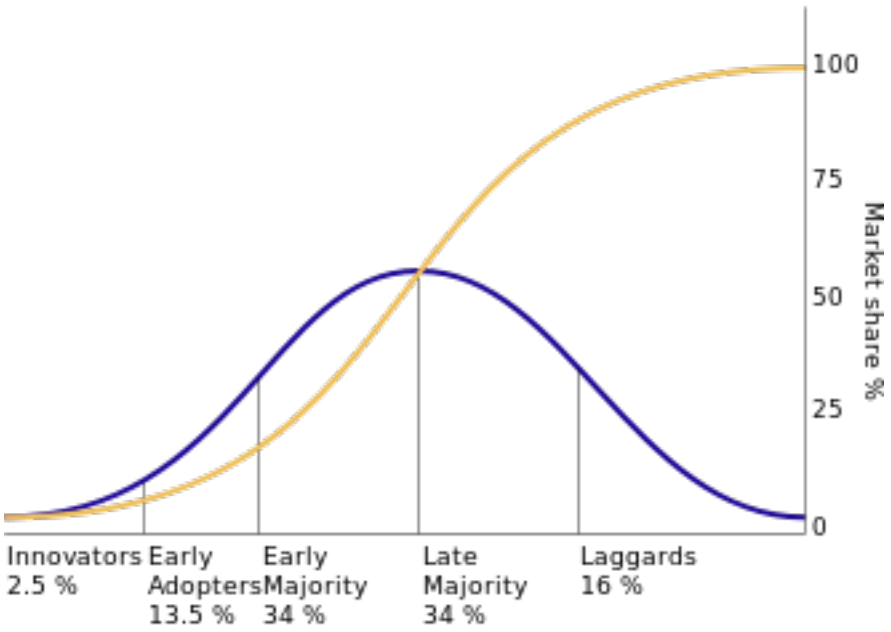

Figure 2.1: The Diffusion of Innovation according to Rogers

.

verification, traverses.

2.1.2 The Hype Cycle Research Methodology

The IT consultancy company Gartner Inc. developed a methodology that took the concept of public visibility further and combined it with the relevance for the market. The model describes the expectations compared to the maturity of a certain technology on basis of the visibility. Gartner Inc. states that the Hype Cycle research methodology "provide[s] a graphic representation of the maturity and adoption of technologies and applications, and how they are potentially relevant to solving real business problems and exploiting new opportunities" (Gartner Inc., 2015). Haklay (2013) sees it as "a way to consider the way technologies are being adopted in a world of rapid communication and inflated expectations from technologies." Technologies are being classified according to the estimated length of time that it may take the technology to reach the plateau of productivity.

According to the Hype Cycle methodology, the media coverage of a new technology goes

through five distinct phases.

1. Technology Trigger
2. Peak of Inflated Expectations
3. Trough of Disillusionment
4. Slope of Enlightenment
5. Plateau of Productivity

Figure 2.2 shows the five different phases depending on the expected maturity of the product to succeed as a business and how the visibility of a respective technology in terms of public discussion, developed products, and business success stories relates to the maturity over time.

Figure 2.2: The Hype Cycle Research Methodology (General Hype Cycle, n.d.)

A technology usually traverses along the line graph and passes several or even all phases in varying speeds. A "Technology Trigger" describes a potential technology or early proof-of-concept which causes interest in the business world and influences the development of products. Visibility is rising, but usability is still limited or even unproven. According to Linden and Fenn (2003) a trigger occurs when "breakthrough, public demonstration, product launch or other event generates significant press and industry interest"(p. 4). An

example may be the explosion in formal verification tools in the late 1990s, when business discovered that failures like the Pentium FDIV calculation may not only result in some minor miscalculation, but also in real money loss.[11]

The next phase is called "Peak of Inflated Expectations," the phase where early adopters take action and spread the word. The proof-of-concept results in the first success stories heating the public discussion, but also is "often accompanied by scores of failures" (Gartner Inc., 2015).

As experimentation and implementation fail to produce marketable results, the technology glides into the so-called "Trough of Disillusionment." This phase is crucial for the technology to survive, as early adopters need to see and feel the benefits of continuous improvements, to not lose interest in the investment. As a result of these improvements, technology can be better understood by the customer, which leads to raising demand and the market entry of second- and third-generation products. According to Linden and Fenn (2003) "more enterprises fund pilots [and] conservative companies remain cautious." This phase is called the "Slope of Enlightenment" and sounds the bell for the "Plateau of Productivity," where a technology, mature enough to result in mainstream adoption, starts to take off.

With the Gartner model, one can derive the amount of risk associated with an investment in a technology. But this model should not be taken as a sole assumption for evaluating technologies. For example, the World Wide Web "hit relatively minor bumps on the fast track to global ubiquity" (Oremus, 2014). However, with respect to this thesis, the Hype Cycle estimates serve to analyze research impediments and opportunities in guaranteeing software correctness.

Depending on the position of a technology on the Hype Cycle, the Hype Cycle may indicate potential side-effects of several technologies or products that may be relevant to classify the relevance of crowdsourced formal verification in today's business world. O'Leary (2008) argues that "the location of a technology on the hype curve drives what types of research questions we can address, what research data is available and what methodologies

[11] In June 1994, the world discovered that (4195835*3145727)/3145727 = 4195579, which resulted in a mass panic when the error was posted in October.

Figure 2.3: The Hype Cycle for Emerging Trends 2013 (Hype Cycle for emerging trends, 2013)

are available to study a technology and its uses"(p. 244).

The Internet of Things (IOT) is a current example of how a technology trend has the potential to transform industries and its effect on humans' life. IOT is not a technology, rather than a concept. It is a collection of different innovations that, combined, affect future business. The concept reached the peak in 2014 (Hern, 2014), but goes already through its second cycle. In 2013, IOT was expected to reach the plateau of productivity in five to 10 years. Only one year later, the new Hype Cycle proclaimed a time period of 2-5 years. However, singular technologies that are pooled under the concept be even closer to the plateau, like the Electronic Product Code (EPC) network with its Radio-frequency identification (RFID) tags for logistics and transportation. These technologies are already comprehensively standardized (Guinard, 2011), while the IOT with all its facets still lacks standardization (Hern, 2014).

Gartner's 2014 Hype Cycle report (Burton & Willis, 2014) identified trends like data science, in-memory database management systems, in-memory analytics, and (hybrid) cloud

computing, that are likely to mature within two to five years, while gamification and big data may reach the so-called Plateau of Productivity within five to 10 years (Figure 2.4).

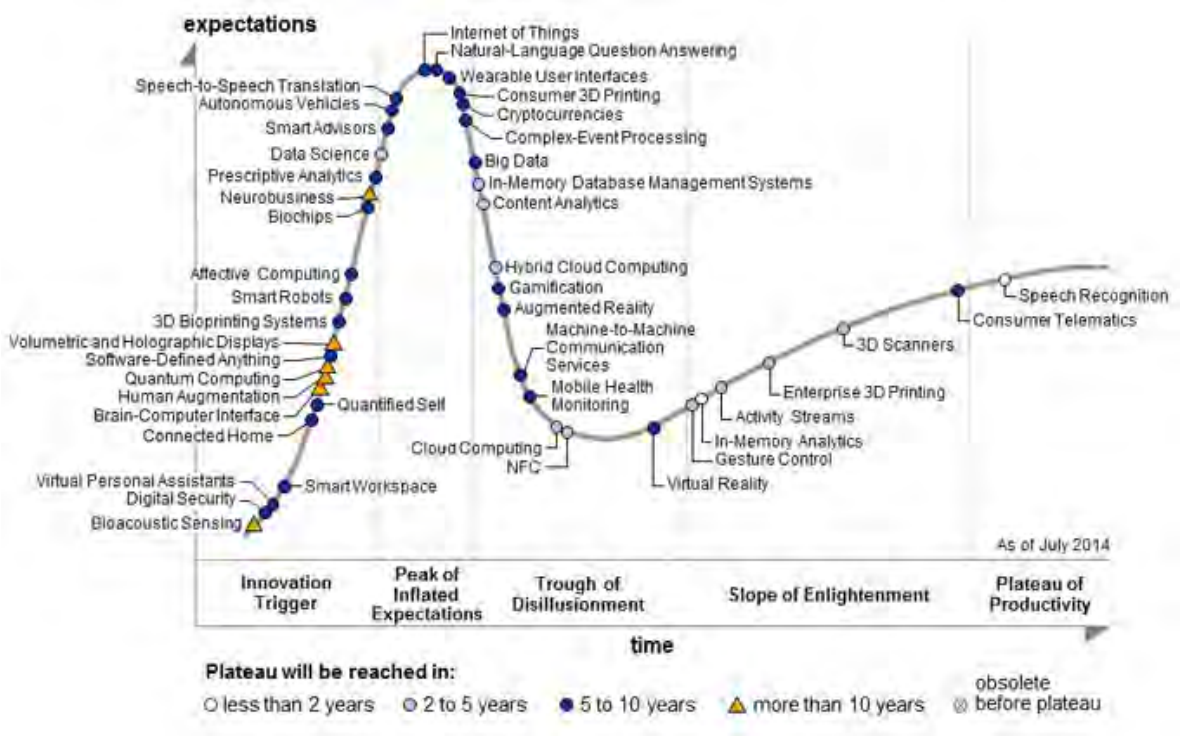

Figure 2.4: The Hype Cycle for Emerging Trends 2014 (Hype Cycle for emerging trends, 2014)

We argue that these trends may have impact on formal verification in the mid-term. Google Trends reveals that the DARPA CSFV project is currently not well known/visible (Figure 2.5). As crowdsourced formal verification of software is a merger of different ideas, it is dependent on the market expectations for these ideas. We reference on these ideas while discussing the current state of development of crowdsourced formal verification, because they either impact the technological maturity or shift the speed and direction of market maturity for crowdsourced formal software verification.

[12]The search criteria were "worldwide, 2004-today, all categories, web search." According to Google, Numbers represent search interest relative to the highest point on the chart. If at most 10% of searches for the given region and time frame were for a specific search term, Google Trends considers this the highest point on the y-axis. It is important to notice that this data does not convey absolute search volume.

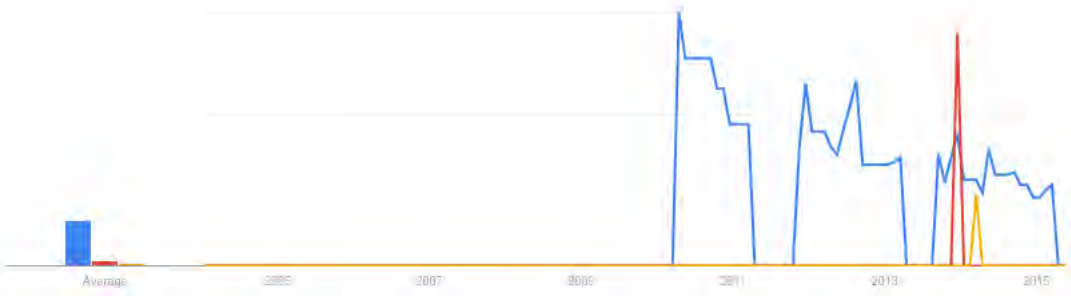

Figure 2.5: The Results of Google Trends for the Search Terms CSFV (blue), Verigames (red), Formal Software Verification (yellow) as of May 2015(Google Trends, 2015) [12]

2.2 Crowdsourcing—Leveraging the Hive Mind

Crowdsourcing is tool of rising popularity that allows business to utilize relatively inexpensive labor provided by people connected through networks.[13] The concept that was first coined in 2006 by Howe (2006) as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call." Yuen, King, and Leung (2011) framed it as "a distributed problem-solving and business production model" (p. 766). Gartner added the concept to its Emerging Technologies Hype Cycle in 2012. Back than, it was estimated to reach the plateau of productivity in five to 10 years. Probably due to its diffusion into other markets, crowdsourcing was removed from the Emerging Technologies Hype Cycle, and added to the Digital Market and Social Software Hype Cycle in 2013.

The widespread accessibility and availability of the Internet allowed to utilize the concept and find new application forms. According to Grier (2013), crowdsourcing can be distinguished by its usage into crowd contests, macro tasks, micro tasks, crowdfunding, and self-organized crowds. Crowd-contests enable to identify the best worker for a specific job, while self-organized crowds are used to embrace competition by offering challenges on the Internet. The Topcoder Inc. community, now part of the Appirio Inc., is one example of a successful platform that gathers design, development, and data science experts by offering challenges and tasks for money. They offer macro tasks that are used to find a specific skill set for a particular job (e.g., web design), or micro tasks if a big job makes it necessary to

[13] Crowdsourcing is a composite of "crowd" and "outsourcing."

split the big job in small pieces. Crowdfunding follows a different approach and tries to fundraise money to startup business as an alternative to classic financing models.

An emerging trend in crowdsourcing, gaining more recognition and relevance over the last decade (cf. Figure 2.6), is the so-called citizen science. A pioneering project was SETI@Home by the University of California, Berkeley in 1999, which has harnessed the idle computing time of millions of participants in the search for extraterrestrial life. Other Internet-based projects followed.[14] In citizen science, people with no formal training contribute to research by collecting data using the scientific method, under the mentorship or supervision of a scientist. Citizen science helps gathering data in an unexpected scale. The Oxford English Dictionary defines citizen science as "scientific work undertaken by members of the general public, often in collaboration with or under the direction of professional scientists and scientific institutions public engagement in collecting" (Citizen, 2015). The two important attributes for this thesis are the one of collaborative research and the purpose of generating new science-based knowledge, the two areas where the Verigames try to gain opportunities from collective collaborative engagement in human-based computation. Wiggins and Crowston (2011) identified five types of citizen science projects.[15] The DARPA games fall under the virtual category. Virtual projects allow the citizens to investigate an issue like real scientist, but they "represent a project type that has not been examined in prior typologies of citizen science" (Wiggins & Crowston, 2011, p. 7), and which use the capabilities and advantages of advanced technologies and gamification means to encourage the people. But they also argue that the tasks that can be done are limited and require extensive web-platforms.

One billion smartphones and 70 million wearable health trackers sold per year opens a new dimension for citizen science (Pogue, 2015). With the launch of the Apple ResearchKit, and the selling of the Apple Watch, citizen science reached a new level. According to a Twitter message from an employee of Sage Bionetworks, a not-for-profit research organization, more than 7400 volunteers enrolled within 6 hours of launch of their App "Parkinson mPower study app" while a conventional research project of a similar scope with great

[14] E.g., Einstein@Home analyses data from gravitational wave detectors, MilkyWay@Home simulates galactic evolution, and LHC@home studies accelerator beam dynamics.

[15] Wiggins and Crowston (2011) distinguished the following types: action-oriented, conversation, investigation, virtual, and education projects.

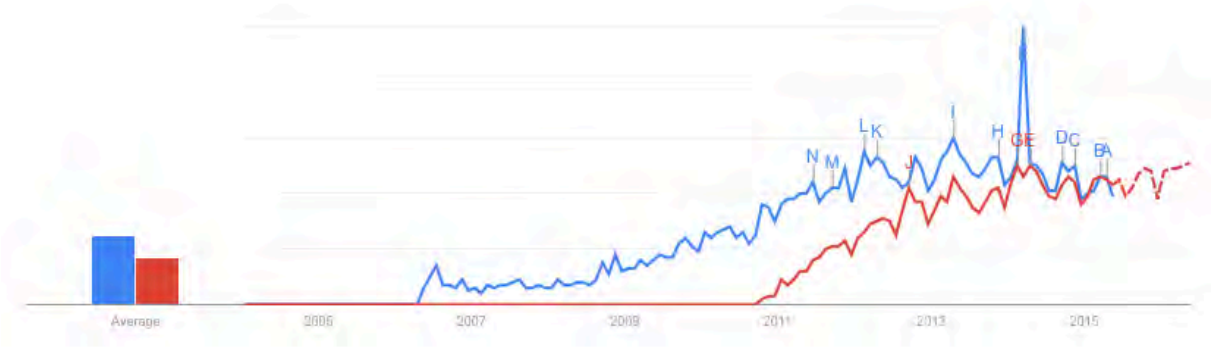

Figure 2.6: Google Trends 'Crowdsourcing' (blue) and 'Citizen Science' (red) as of May 2015 (Google Trends Citizen Science, 2015)

difficulty usually would involve around 1700 people (Sreeraman, 2015). However, the retention rate in citizen science is very low. According to a study on citizen science projects by Sauermann and Franzoni (2015) the most compelling projects attracted only 40% to contribute a second time. The average visit time ranged from seven minutes up to 25 minutes. In contrast, the cost savings for the programs turned out to be the key driver for success. According to the study, even the project with the fewest participants saved $22,000, while all seven surveyed projects saved about $1.5 million.

2.3 Gamification

Keeping the attention of customers or users is one of the difficult tasks for successful businesses. One means to gain attention and create engagement by the users is game design, borrowing elements from video games. According to Marczewski (2012) the concept was first coined by Nick Pelling. The concept did not raise widespread popularity until around 2008 (Deterding, Khaled, Nacke, & Dixon, 2011). In 2011, the Gartner Inc. added the concept to its hype cycle chart (Dorling & McCaffery, 2012) and predicted that in 2015, over 50% of organizations that manage innovation processes would gamify those processes, and and more than 70% of Global 2000 organizations will have at least one gamified application (Gartner Inc., 2011).

Enterprises today use badges, rewards, and other elements from video games to accelerate their business and bond long-term customers. There are several different definitions of

gamification discussed in academia. A commonly cited definition referenced on the Internet was developed by Zichermann and Cunningham (2011), who defined gamification as "the process of game-thinking and game mechanics to engage users and solve problems."[16] They argue that gamification compiles different concepts that have been advanced in games for a non-game context to influence behavior, and to create engagement and loyalty through rewards. Also in 2011, Deterding et al. (2011) proposed the academic discussion and referred to gamification as "the use of game design elements in non-game contexts." These elements can encompass narrative guides and challenges, rapid indication of success, support and competition factors, and aesthetical design considerations. Deterding et al. (2011) argue that the concept of gamification should not be limited to specific usage contexts, purposes, or scenarios, while "engagement, or more generally improving the user experience serve as popular usage contexts." According to Deterding et al. (2011), gamification is more a design element with a wide variety that can be applied to different domains.

Today, the concept of gamification is applied to almost every business domain. For example, businesses added game mechanics to educate their personnel, or to bond customers or users. The market is expected to grow by 99% between 2012-2016. The website www.enterprise-gamification.com lists over 80 examples of enterprises that reported the successful implementation of game design elements in a non-game context (cf. Enterprise Gamification Consultancy, 2015). Although the applied concept only exists for a short period of time, some of the early adopters already decided to get out of gamification. One of the most successful examples of applied gamification is Foursquare, a local search and discovery mobile application, from Foursquare Lab Inc. The app developers reported in early 2013 that they would remove game design elements from their service, as it was not aligned to their business goals anymore (Hepp, 2013). The decision showed that a successfully applied concept may require re-evaluation if the business model changes. But gamification can also have immediate impact on the business success. In 2013, Comcast, a U.S. cable TV and Internet provider, applied gamification tactics to its sales representatives. In opposite of the former spreadsheet and report driven sales competition, the embracing of a

[16]The term game mechanics is widely discussed in the literature. For this thesis, we use the definition by Cook who defined game mechanics as "rule based system/simulations that facilitate and encourage a user to explore and learn the properties of their possibility space through the use of feedback mechanisms" (Cook, 2006). This definition is valuable to the gamification discussion, as it includes the feedback/ reward component.

gamification app led to an increase of 127% in booked appointments (Vehns, 2014).

Crowdsourcing is a major field in which gamification has been applied so far. Gamification can be an additional motivation factor to crowdsourcing initiatives (Barsky, 2012). Most crowdsourcing initiatives focus on a short- to medium-term relationship to the customer, as for example crowdfunding initiatives. But gamification can help to foster the long-term relationship necessary for a continuous success of the business model like "FoldIt" or "Play to Cure: Genes In Space" (getting more research results), "TrashOut" or "Greenify" (receiving continuous input on data from users), or "DIRECTV" (to spur employee development and make everyone equal in terms of ideas to improve processes (Greengard, 2014)). Tellioglu (2014) found that the engagement rate for CSFV games is very low in comparison with traditional games. Gamification has been applied on the Verigames.com portal or the games themselves.

2.4 Formal Verification of Software

Formal Verification is based on several different concepts. In the following subsection, we discuss the terminology used in conjunction with the formal verification of code. We distinguish the terms *validation* and *verification*, and describe how formal methods ameliorate software verification.

2.4.1 Verification and Validation

Verification and validation of software are two methodologies to determine a software's fitness and to assure software safety or correctness. Software engineers use verification and validation to provide usable software code to users. Both terms are often used imprecisely and they are not clearly distinguished.

In this thesis, we rely on the following definitions: *Software validation* ensures that software meets the user's needs, and that it "fulfills its intended use in its intended environment" (Softwaretestingfundamentals, 2011). *Software verification*, on the other hand, determines "whether the products of a given development phase satisfy the conditions imposed at the start of that phase" (Radatz, Geraci, & Katki, n.d.). We focus on the verification aspects only. Easterbrook (2010) recommends thinking about verification and validation as a toolbox that provides "a wide range of tools for asking different kinds of questions about

software."

This offers a valuable perspective, as hardware and software systems continuously become more complex due to increasing scale and functionality. Figure 2.7 gives an overview of the terminology used in verification and validation to provide a better reference for the following explanations.

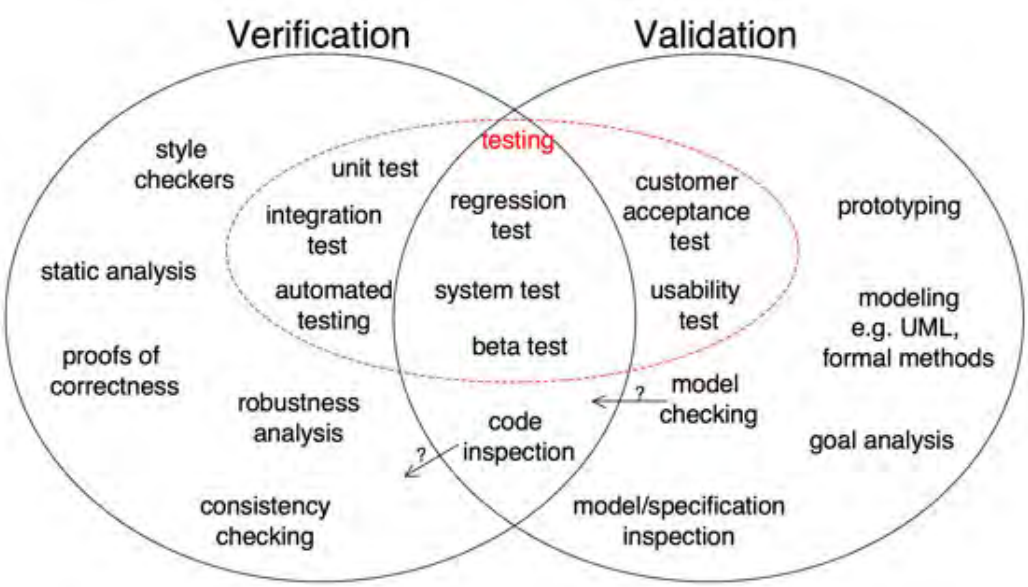

Figure 2.7: The Verification And Validation Toolbox (Easterbrook, 2010)

2.4.2 Formal Methods

Formal methods can reduce the complexity of verifying systems (Clarke & Wing, 1996). By using formal methods, complex models are represented by mathematical entities, which allows specifying, designing, and verifying the system's properties more profoundly than empirical testing (Butler, 2001; Collins, 1998). In software verification, formal methods are important, as they reveal inconsistencies, ambiguities, and incompleteness that might otherwise go undetected by testing (Clarke & Wing, 1996; Kroening & Sharygina, 2005). Therefore, we follow the definition of formal software verification by Li as "an act of using formal methods to check the correctness of intended programs" (Li, 2010). However, we concede that mathematical proofs only provide high assurance that code matches its specifications, but cannot guarantee software correctness and safety.

Formal software verification is still dominated by the labor of human experts, who possess high mathematical understanding and skills. With advancing technology, new tools became available to automatically or semi-automatically verify properties of a computer program during the compilation or run-time, or based on analysis of a program's text (source code). Consequently, Logas et al., the authors of the CSFV game Xylem, defined formal software verification as "the use of tools such as proof assistants and model checkers to automatically or semi-automatically verify properties of a computer program under consideration" (Logas et al., 2014). According to Yuen et al. (2011) these formal tools are in widespread use providing verification of source code based on different implementation techniques like abstract interpretation, automatic theorem proofing, model checking, and type checking. They also found out in their survey that more companies are planning to adopt and use formal verification. Unfortunately, these tools are still limited by the complexity, size, faster development of code, and middleware. Sa'ar (2010) and Ray (2005) identified model checking and deductive verification as the most reasonable for automated verification.

Model checking proofs whether a program meets its specification based on finite state models. This technique is said to be fairly fast and can be fully automated, but is unsuitable for verification of large models as the number of states grow (state explosion) (Strunk, Aiello, & Knight, 2006). Therefore, finite state models have to be relatively abstract to allow verification. This technique is for example used by the CSFV game FlowJam.

Deductive verification is the basis of the CSFV games Stormbound and Xylem. Deductive verification "expresses the correctness of a program by a set of mathematical statements, called verification conditions" (Filliâtre, 2011). These verification conditions are symbolic representations, written in logical languages, often in a first-order logic, or higher-order-logic (Jaffar, Murali, Navas, & Santosa, 2012; Schellhorn, Wehrheim, & Derrick, 2012). They can be proofed automatically or interactively (semi-automatically) by theorem proofers, which work sequentially based on primitive recursive functions, resulting in scalability problems. These theorem provers demand the help of human experts as computer algorithms are still not mature enough to independently decide whether a specification is met. In summary, formal verification is useful for finding certain bugs. It depends on a verifier and the description of specification of software code. If successful, we get a verified system that, with respect to the given properties, is mathematically guaranteed to be correct.

CHAPTER 3:
Problem Diagnosis — Analyzing Crowdsourced Formal Verification

> Today a usual technique is to make a program and then to test it. But: program testing can be a very effective way to show the presence of bugs, but is hopelessly inadequate for showing their absence.
>
> —Edsger Dijkstra, Notes On Structured Programming, 1970

This chapter diagnoses the current problem areas of the CSFV approach by the example of the game Stormbound. We discuss whether time will take care of the identified issues and whether there is a solution space, or the CSFV concept is intrinsically flawed.

3.1 Technical Deficiencies of Interactive Formal Verification

We are going to do a quantitative assessment of the back end data from Galois. We try to identify the current potential of returned assertions of the game players, while revealing technical shortfalls of concurrent verification software. The dataset includes 146,595 valid assertions successful solutions generated by the players in over 2,465 hours. The Stormbound Team reported that players had contributed to 4,361 out of 6,523 levels, which is 68% of the games levels (DARPA, n.d.).[17]

3.1.1 The Formal Verification Process behind Stormbound

This subsection describes the Stormbound specific way of formal verification. This process is unique and does not describe the way how the other Verigames work in the background. However, it will help us to gain a better understanding when analyzing the data. We first

[17] Because every player may come up with unique assertions, levels can have multiple solutions.
In a working paper, the Stormbound team claims 142,711 valid assertions that are generated by players in over 2,919.2 hours (excluding CSFV team members) (DARPA, n.d.).

describe the three phases of the Galois approach to follow up with a more detailed perspective on what assertions look like. The following process description was developed through extensive discussion and written conversation with the Galois representatives and a review of literature about FRAMA-C (Cuoq et al., 2012; Könighofer, 2013).[18]

The phases of the Stormbound Verification Process

The main goal of the Stormbound game is to solve verification conditions, generated by FRAMA-C, a set of interoperable program analyzers for C programs. Stormbound approaches the formal verification of software in three consecutive phases.

In the first phase, a FRAMA-C plugin, called Runtime Error (RTE)s, checks the BIND code used in the CSFV program for common Runtime Errors. These RTEs are correctness properties of a program that are obvious, such as a divisions by zero, or reading or writing invalid memory locations (the typical array-index out-of-bounds problems). The plugin inserts annotations for these RTEs automatically. Afterwards, FRAMA-C compiles with special user supplied instrumentation routines additional assertions for pre- and post-conditions, and invariants. For instance, if a program contains loops, then Hoare Logic requires loop invariants for the proof. Because only some loop invariants are easy to find, the user has to define the ones that are difficult to find automatically (Könighofer, 2013). The produced assertions for RTEs, pre-, post-conditions, and invariants are the proof obligations, so-called goals, or proof goals. The Galois team then produces snapshots due to the large amount of data, which would cause performance constraints. Snapshotting the entire heap is too slow and produces too much data. These snapshots They are the basis for assertions. They are then post-processed to handle some tedious C issues like memory leaks, and are being sent to the game database.

The second phase involves embedding some of these snapshots for a single program point in a Stormbound game instance. While playing, the player guesses an assertion and wins when the check by the back-end returns that the assertion is true.

The final phase involves running FRAMA-C to get the holes filled by sending the user assertions to a theorem prover back-end. The used theorem prover is Alt-Ergo, which is based on Satisfiability Modulo Theories (SMT), which attempts to verify that the constructed

[18] S. Winwood, personal communication, 05 and 10 March 2015.

assertions from user play are sufficient to prove the goals.

Details on the Stormbound Verification Process

Verification conditions are added by FRAMA-C whenever there is an unsafe operation, which would cause a run-time error such as memory access or potential overflow. We generally use the term RTE to refer to these assertions.

FRAMA-C produces a number of goals that may assist to prove these RTEs will never occur. This process is not fully automatic. It requires input from a verification engineer. In particular, known invariants and skeleton definitions for potential invariants need to be added to loops, while pre- and post-conditions are added to functions. These extra assertions are called "holes" as they need to be filled for the verification to work. The objective of the Stormbound game is to fill these holes with additional assertions such that the goals are proven true.

The approach so far produced several sorts of goals: those which relate directly to RTEs, and those which show that the other assertions (invariants and pre/post conditions) are true. For instance, given the following loop invariant in Figure 3.1 we have pre-condition P, loop invariant I, and post-condition Q. We seek to come up with assertions for these such that the memory access, in this case the *x > 0 is always safe, and so that the assertions are true.

```
1 {P}
2 void foo(int *x)
3 {
4         while(*x > 0) {invariant I}{
5                 x++;
6                 }
7         *x = 1;
8 }
9 {Q}
```

Figure 3.1: Example of a Loop

The goals for this will then be something like Figure 3.2 on the following page, which means that we get four goals, from this one function.

```
assuming P, show that I holds
assuming I, show that *x is OK
assuming I, show that I holds after the body of the loop
assuming I, show that Q holds after updating *x
```

Figure 3.2: Goals for the Function in Figure 3.1

If there is a function call, then there is an additional class of goal which is that the pre-condition of the called function holds. For example, if we had instead Figure 3.3 then we would have the additional goal assuming I, show that the pre-condition of f holds.

```
while(*x > 0) {invariant I}{
        x++;
    f(x);
}
```

Figure 3.3: Goal for a Function Call

Contribution of the Game Play

Users get a single program point to guess for per game play. The Stormbound game shows users the values of various variables at these holes, i.e., the snapshots generated with the help of Frama-C. The users are asked to guess assertions which are true at that point. In an ideal world (having many players), the players generate many user assertions per hole, which together form the final hole assignment. For example, if the players produced a1, a2, and a3 for I, these user assertions will be filled in I as a1 AND a2 AND a3. This means that the Stormbound game produces only aggregate data. User-submitted assertions need to be verified in combination, not in isolation, as, for example, it needs to be tested that the invariant holds assuming the pre-condition, both of which are holes, and both of which come from user play, notably unrelated user play.

As a consequence, any percentage calculated for proofed goals in the next subchapter cannot be considered as "final" coverage of verified code.

3.1.2 Quantitative Analysis of the Games Data

The quantitative data from the Stormbound game was provided by the team of Galois, specialists in formal methods, and voidALPHA, a video-game studio. The team provided game session data, returned assertions of the player, and results of the proofer.

Data Preparation

The dataset consisted of three tables and two source code files. The formats of the CSV and TXT files are as described in the following figures 3.4, 3.5, 3.6 on the next page. The source code files contained the original BIND code and the annotated BIND code from FRAMA-C.

One CSV file contained session times as recorded by the game client.

```
kwSiipbJdrxkNkpG,0,,Wed Oct 23 2013 17:06:53 GMT-0400 (EDT),154\\
(user id, ignore, ignore, start time, session time in seconds)
```

Figure 3.4: Format of the CSV File Containing Session Time

One TXT file contained the number of player-produced assertions for each program point ("Hole"), with P for precondition, Q for post-condition, and I<n> for each invariant in that function.

```
add_trace_entry I1: 257 I2: 20 I3: 92 P: 17 Q: 14\\
(function name, first invariant, second, third, precondition,
postcondition)
```

Figure 3.5: Format of the First TXT File Containing Assertion Counts for Each Program Point

A second TXT file contained results from running the checker on the submitted assertions. In this case, the first invariant has 257 distinct assertions, the second has 20, the third has 92. The players also produced 17 assertions for the pre-condition, and 14 assertions for the post-condition.

```
1 add_trace_entry RTEs: 9/19 Calls: 2/2 Loops: 1/4 Post: 1/1
2 (function name, rte result, function call result, loop result,
3 post-condition result)
```

Figure 3.6: Format of the Second TXT File Containing Results from Running the Checker on the Submitted Assertions

The results for RTEs, Function Calls[19] , Loops, and Pre-Conditions are of the form "n/m," where "n" being the number of goals which are proven, and where "m" being the number of goals.

To analyze the data, the file content was imported into Microsoft Excel. Due to the different amounts of columns of assertion counts for invariants, the data needed to be reorganized. As the players did not submit assertions for all functions or some functions had been already solved by automated tools, the count of functions in both tables was different. Therefore, the data of assertion counts and the results of the checker were concatenated by using MS Excel's consolidation function. This revealed insights in the shortfalls of FRAMA-C or potential bugs in the process. Some of the functions had no assigned goals, but players submitted assertions for these functions.

Some functions showed goals were listed for the functions, but assertions had been submitted by the players.

Galois also provided the *post-source.c* file, which contained the annotations of the FRAMA-C plugin and the Holes added in the first phase. Based on this file, we want to reiterate the data that the players produced are individual assertions, something like (as a completely made up example):

$$\boxed{x > 10 \text{ or } y = 5}$$

These assertions are for a single hole (pre-, post-condition, or loop invariant). All submitted assertions that are produced for a hole are taken to come up with the final assignment. Ideally we would come up with the smallest useful set of assertions for each hole, but

[19]Note that the Function Calls serve as Pre-Conditions.

figuring out what this actually is would be computationally infeasible (or would at least take too long).

Data results

Between December 2013 and September 2014, the Stormbound players generated 146,595 assertions for 1,958 functions. The results are summarized in Table 3.1. The table lists the maximum amount of pre-conditions, invariants, or post-conditions for the holes of a single function. The "Avg" column shows the number of assertions per pre-conditions, invariants, or post-conditions for the holes of a single function were produced per function on average.[20]

	Sum	Min	Max	Avg
Pre-conditions	63,187	0	2,477	32
Invariants	27,356	0	7,712	29
Post-conditions	56,052	0	1,268	14
$\sum$	146,595			

Table 3.1: Assertions Produced by the Players

The data also provided results for 4,473 functions of the BIND code. FRAMA-C produced goals for 3,810 functions. We found several anomalies that let us assume that the dataset is limited for finally assessing the verification success of the Stormbound game.

First limitation: For 663 functions, FRAMA-C did not calculate respective goals. This can be caused by different reasons. There are some functions, which we do not have results for either,

- because they are not called by the program (so-called dead code),
- because there is some error with running the FRAMA-C test-tool suite on them,
- or because they produce so many goals that processing them is not feasible.

Second limitation: The players generated assertions for 1,958 functions. However, 154 assertions were generated for functions without any goals, which should not happen. If no goals are calculated by FRAMA-C, no assertions should be produced by players.

[20] As mentioned in Section 3.1.1 on page 28, Stormbound only generates aggregated data, and there is no 1:1 connection between goals and holes. Therefore, we cannot directly link the produced assertions to goals, which means one cannot infer the necessary amount of assertions produced per function.

Third limitation: We assume that goals are relevant for players contributing to the verification process. Due to the fact that only for 1,804 out of 3,810 functions goals were generated, the maximum function level coverage of verified by the game players would be 47.35%.[21]

Fourth limitation: Some verification runs of the back-end encountered errors, so the results for these are not included.

The found data is summarized in Table 3.2.

Functions	# of Functions with assertions	# of Functions without assertions	∑ of Functions
with goals	1,804	2,006	3,810
without goals	154	509	663
∑	1,958	2,515	4,473

Table 3.2: Assertions vs. Functions

For the further evaluation, we omitted the functions without goals assigned by FRAMA-C. An optimal result would show a equal number of goals and proofed goals. For example, we found that 47,508 goals were generated for the RTE/Holes, but only 23,759 of these goals were proofed. We calculated different percentages. First, as the ratio of total numbers of goals proven to the total number of goals (e.g., the sum of goals, which are proven for RTE divided by the sum of RTE goals for all functions). We call these ratio "Proofed goals on Totals." Second, the average of all ratios for the single functions with RTE goals. We call that ratio "Proofed Goals on Functions." Unfortunately, this distinction does not provide a significant insight.

Goals for Run-time Errors: For 3,063 functions RTEs were proofed. The following Cumulative Distribution Function (CDF) describes the statistical probability of the proofed RTE goals(Figure A.2). It shows that approximately 22.7% of the functions' RTE goals were not solved. 18.24% of the functions' RTE goals were fully verified. For the rest of the functions, partial RTE goals were proven.

[21]The system may generate multiple goals for a single RTE. Also, if there are no user assertions then the prover assigns true for a hole. This makes it simple to prove that a loop invariant was preserved.

	# of Goals	# of Proofed Goals	Proofed Goals on Totals [%]	Proofed Goals on Functions [%]
RTEs	47,508	23,759	50.01	45.49
Pre-Conditions	46,327	38,337	82.75	90.44
Loop Invariants	3,058	2,303	75.31	76.20
Post-conditions	6,644	5,445	81.95	79.66
Total Goals	103,537	69,844	67.46	72.95

Table 3.3: Comparison of Goals vs. Proofed Goals

Goals for Pre-Conditions: The players' assertions helped to prove pre-condition goals for 3,497 functions. The CDF (Figure A.3) shows that almost all function call goals were solved. Approximately 76% of the functions call goals were fully verified. Only 1.6% of the goals were not proven.

Goals for Loop Invariants: For 763 functions, Loop Invariant goals were found. Figure A.4 on page 66 shows that loop goals of 10.62% of the functions were not proven. 61.6% of the functions' loop goals were fully verified.

Goals for Post-Conditions: For 3791 functions post-conditions were proven. We can see that 19.84% of the functions' post-condition goals were not solved. 79.03% of the functions' post-condition goals were fully verified. Figure A.5) shows an odd behavior of the curve. The curve is almost horizontal between the above mentioned values. This results from only 1.13% of the functions having partial goals proven. Almost all functions had no or only one post-condition goal. This small fraction, however, had between 2 and 971 post-condition goals.

A Consolidated Picture: The consolidated picture in Figure A.6 showed that only 3 functions out of 3,497 have zero goals proven. Also, only 25.08% of the functions have all their goals proven. The players' assertions helped to solve only a minority of the verification goals.

3.2 User Participation in the Stormbound Game

In the following subsection, we examine published articles, usability test results, and lessons learned of the different teams. We can summarize that the front-ends are generally mature, although they need to provide a more sophisticated environment for user interactions and

rewards. Although the games gained high visibility during the first month of release, the lack of constant player numbers was a common issue for all the games (Tellioglu, 2014) and the engagement rate did not improve over time. Some games like Stormbound found the need to increase their level of mathematics in the games, while other games needed to reduce the requirements on the user's mathematical skills.

3.2.1 Player Contribution

In a first step, we examined the performance of the Stormbound game based on the data for the player's sessions of phase 1.

During that time, 14,964 distinct players played the game over 524 days. We found that the minimum and maximum values for players per day stayed the same (Table 3.4) compared to the examination in 2014 (cf. Tellioglu, 2014) for the whole phase 1. The daily participation span from 1 on eight days to over 800 on three days.

Sum	Min	Max	Mean	Median	StDev	Kurtosis	Skewness
14,964	1	860	28.56	12	90.03	56.1	7.23

Table 3.4: Descriptive Statistics for Stormbound Players during Phase 1

The player base stayed constantly below 50 participants daily after May 2014 (Table 3.5). Due to the low participation, the mean and median decreased drastically (mean: 71.5 -> 28.56, median: 23 -> 12).[22] The high value for kurtosis shows that the distribution had a high peak (here: December 2013), and the positive skewness proves the long-tail distribution to the right (Figure A.1 in the Appendix), which means that the hype was at the early beginning of the phase with some minor peaks in January, April, and May 2014.

Table 3.5 shows that there was no constant monthly average user base during phase 1. The player numbers changed from month to month significantly.

3.2.2 Stormbound's Engagement Rate

Tellioglu calculated on the basis of the available data from December 2013 till May 2014 the metrics daily average user and monthly average user, and then derived the engagement rate. He found that the engagement rate of the players for the Verigames, the so-called

[22] We disregard May 2015 in this statement, as the data was only provided for nine days of the month, which bias the results.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
’13												
Px												7,552
Δ%												
’14												
PX	956	617	412	464	532	594	327	376	239	339	293	236
Δ%	-87.3	-35.5	-33.2	12.6	14.7	11.7	-44.9	15.0	-36.4	41.8	-13.6	-19.5
’15												
Px	170	185	175	171	29							
Δ%	-28.0	8.8	-5.4	-2.3								

Table 3.5: Monthly Player Numbers of Stormbound during Phase 1

stickiness, is below the industry standard, which is between 10 and 30% (Tellioglu, 2014), and that the games have a high drop-off rate.

Days	Min [%]	Max [%]	Mean [%]	Median [%]	StDev [%]	Kurtosis	Skewness
514	0.05	15.99	3.66	3.40	2.13	8.37	2.19

Table 3.6: Descriptive Statistics for Stormbound’s ER during Phase 1

We reviewed this data for the whole phase 1 from December 2013 till May 2015. Table 3.7 provides an overview of the monthly engagement rate of Stormbound. We found that the daily ER was not stable over the whole time. Over 514 days, the game had several peaks in the daily ER. The distribution in Figure A.7 in the Appendix shows a dominating low ER on most days over the phase. Figure A.8 shows the daily average ER, while Figure A.9 shows the monthly average ER and participation during phase 1.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
’13												
ER												3.43
’14												
ER	3.86	4.46	3.93	3.73	3.75	3.73	3.58	3.53	3.50	3.43	3.63	3.51
’15												
ER	3.30	3.69	3.49	3.67								

Table 3.7: Monthly ER of Stormbound during Phase 1.

We assumed that the player ER would be dependent on the time of the year. Using regres-

sion analysis, we found that the Engagement Rate was not dependent on the time of the year. With a 95% confidence level, the p-value for year was 0.195530015. The p-value for month was 0.084042062, and the R^2 value was 0.006015543. As the p-value was higher than 0.05, we had to reject the null hypothesis. Hence there was no significant relationship between the variables in the linear regression model of the data set. Accordingly, the respective R^2 value shows that the variables month and year cannot explain the ER.

3.2.3 The Problem With the Niche

Formel methods are based on mathematical methods. Two of the games, Xylem: The Code of Plants and Stormbound, went different ways to deal with the necessary math skills of players to contribute to the formal modeling of a software system. In Xylem, players depend on mathematical observations about synthetic plants to solve the games. Also while the developers of Xylem wanted to "soften the emphasis on math" on the one hand, they had to provide a certain amount of math to help the players find a reasonable amount of loop invariants (Logas et al., 2014). According to the team, the final product emphasized more math than a "casual audience would be comfortable with" (Logas et al., 2014).

While Xylem tried to find its niche with a casual-math customer base, Stormbound chose to not confront the players with math at all and completely hid the math. Stormbound allowed players to make assertions without any math or numbers in-game.

However, both approaches had to weight between the amount of players, and the success of generating user assertions. As lessons learned from the phase 1, both games decided to adjust directions. This time, Stormbound wanted to "give players tons of context, and focus on efficiency and comprehension" (DARPA, n.d.). The Stormbound team assumed that the players are much more math savvy and would, similar to citizen scientists, be interested in the underlaying mathematical constructs of the games. At the same time, the Xylem team found that their approach attracted a much smaller audience than they needed to formally verify the software in the background, and that their top 20 players would be heavily interested in math (DARPA, n.d.).

For phase 2, both teams decided to shift the focus on a citizen scientists who would be interested in cybersecurity, and who would have the necessary mathematical skills to contribute to a more math-intensive game.

3.3 Conclusion on the Data Analysis

A fully automated approach is still far away from being realized, unless the Rice theorem, which states that any non-trivial property of programs is undecidable, can be disproved. The seL4 kernel verification already showed that automatization is helpful (Klein et al., 2010). The above shown results stand exemplarily for one game out of five games of the Verigames. No general conclusion can be drawn to the overall functionality of a crowd-sourced human-assisted approach in formal verification of software. Too different are the approaches of the other Verigames of the CSFV phase 1.

However, the data shows that with the first attempt only 25.08% of the functions had all goals proven. Less than 50% of the RTE identified by FRAMA-C were proved. The contribution of the players only allowed to find solutions for less than 80% of the goals of the pre-conditions, invariants, and post-conditions that would determine a formal verification.

Also, the player analysis confirmed the results from 2014. Stormbound could not attract more players over time. The engagement rate continued to be low. Seasonal trends could not be identified. Using the Pareto rule would lead us to the conclusion that no predictions can be made so far about the real effort and participation of non-experts necessary to produce enough user assertions to fill the holes and prove the goals. Therefore, the results of the Stormbound game do not allow a sound verification statement so far. However, the data shows that crowdsourced formal verification of software is at least a contribution to formal verification of software, as it allows non experts to participate in the finding process of assertions, which otherwise would be a closed community of formal verification experts limited in numbers by its sheer head count compared to the daily growing amount of code.

Interesting are the assumptions of both teams for the second phase. Based on the work of Tellioglu (2014), the teams decided to lever the curve for math in the games. This will further limit the potential user base and is most likely not leading towards permanent higher player numbers. Moreover, it will push the games into a niche, that may limit its attractiveness.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 4:
Future Prospects on Formal Verification

> Digital technologies had been laughably bad at a lot of things for a long time - then they suddenly got very good.
>
> —Erik Brynjolfsson & Andrew McAfee, The second machine age: Work, progress, and prosperity in a time of brilliant technologies, 2014

In this chapter, we look closer at trends in gamification and crowdsourcing, especially citizen science, the so-called crowd-sourced science, to explore whether such an approach is expedient. Based on the findings, we discuss how these innovations can contribute to the maturity of the crowdsourced formal verification of software.

4.1 Expectations on Gamification

Gamification is a trend that continuously emerged over the last decade to a well-recognized business factor in different domains. Workman (2013) goes so far as to say that gamification "represents the fusion of four trends: the explosion of social media usage, the mobile revolution, the rise of big data, and the emergence of wearable computing." According to 53% of the participants during the 2014 survey by the Pew Research Centers Internet & American Life Project and Elon University's Imagining the Internet Center concluded ("Imagining the Internet," 2015) gamification will be widespread.[23] Bing Gordon, partner at Kleiner Perkins, a venture capital firm located in Silicon Valley, claimed in 2011 that "every startup CEO should understand gamification, because the gaming is the new normal" (Tsotsis, 2011).

During 2014, applied gamification leveled up and revenue numbers followed the prediction of the M2 Research advisory group, focused on the convergence of digital media,

[23] 1,021 technology stakeholders and critics responded to the online, opt-in survey. 53% said that gamification will be widespread, but a number of them qualified this by saying the evolving adoption of gamification will continue to have some limits. 42% chose a more modest scenario that predicted gamification will not evolve to be a larger trend except in specific realms ("Imagining the Internet," 2015).

entertainment and business applications, who expect continuous growth of the worldwide market for gamification up to $2.8 billion in 2016 (M2 Research, 2015). BI Intelligence, a subscription-based syndicated research and information service, even predicts the trend moving up till 2018, reaching $5.5 billion by 2018 (Workman, 2013). Figure 4.1 shows the market forecast extrapolated starting 2011.

Figure 4.1: The Gamification Market Forecast by BI Intelligence (Workman, 2013)

Although the concept has already been well received in the business world, gamification has entered the trough of disillusionment (Burton & Willis, 2014). The market starts to better understand the issue and public discussion becomes more settled. Workman (2013) calls this the "demise of superficial gamification," which means that the market starts to have more sophisticated demand on the application of gamification. Virtual badges and intangible rewards are no longer a sole source of customer retention. Also, due to the advances and facets that the concept today shows in different business sectors, the term's comprehensive correctness is already questioned (Anderson & Rainie, 2012).

Looking at the discussions in Internet forums, one can identify that there is a common understanding that a "one-size-fits-all" gamification does not meet the user's expectations anymore. Applying gamified design elements to products and services has to provide real recognition to the users in order to to embrace engagement. Gamification cannot be applied out-of-the box anymore, because users cannot be tricked anymore by "earning goofy badges and trophies" (Paharia, 2015). Having get used to the gamification means during the last decade, the users have been come of age, and want to enter in a win-win situation, rather than consuming meaningless game mechanics (Burke, 2013; Paharia, 2015).

Applied gamification that enables the player to gain recognition can be utilized to achieve a change in behavior, developing skills, and enable innovation to meet business objectives (Gartner Inc., 2012; Poser, 2015). Opower, a publicly held Software-as-a-Service company, is an example of how gamification is a subset of these social influence processes. Opower offers cloud-based software that connects utility providers and their customer. Through the use of extensive gamification customers gain awareness of their energy consumption behavior compared to other customers. By getting customized feedback and social comparison, Opower thrives in part on competition as well as amusement. Customers finally win by having reduced utility costs through adapted behavior in energy consumption. Opower applies the principles coined by Cialdini's insight that "people's ability to understand the factors that affect their behavior is surprisingly poor" (Cialdini, Goldstein, , & Martin, 2008).

Following this understanding, a business in crowdsourced formal verification should apply gamification two-folded to meet its business objectives: it should raise awareness, reward personal experience and contribution to the verification in a sophisticated manner, and allow users to explore and experiment their problem solving skills in a collaborative environment. The reward system must be integrated in the platform, and not isolated for single part (e.g., single games). Blaney (2014) argues that an effective gamification platform should combine three components:

1. Applied knowledge of intrinsic motivation,
2. Big data analytics,
3. Scalable and sustainable capabilities.

Comparing the above statements with the Verigames.com platform for phase 2, which started in May 2015, sophisticated gamification still has to be embedded or is in its infancies.[24]

4.2 Crowdsourcing and Citizen Science

Future prospects await projects utilizing citizen science in the private sector. Several projects have already proven success to utilize the wisdom of the crowd to accomplish tasks that computers could not accomplish by embracing volunteer's thinking. The question remains, whether for-profit companies can provide an environment that inspires volunteers to invest their time and effort to sustain projects that benefit shareholders rather than society at large. Surowiecki (2005) found that four criteria need to be met to use this "wisdom of the crowd":

1. Decentralization: different opinions are cumulated utilizing the web
2. Diversity: Different expertise is utilized
3. Independence of individual opinions
4. Aggregation: the contribution of the crowd is homogenized and aggregated for the job accomplishment.

Following these "golden rules," we can conclude that we need to open the CSFV approach to a worldwide community. However, according to Grey (2009), the citizen scientists can be primarily utilized in the industrialized world with the necessary equipment and leisure time. This may hold true for the traditional citizen science projects that focus on environmental of geographic information science, but may be irrelevant for virtual projects like CSFV. Nevertheless, for-profit organizations that want to utilize the scientific-oriented crowd need to provide an environment that allows participants to be recognized as a scientist, because they will openly advocate their contribution (Toerpe, 2013).

Shirk et al. (2012) developed a framework (Figure 4.2 on the facing page) that helps to integrate these efforts into successful outcomes. The framework promulgates the dependencies of public participation in scientific research projects and provides a story plot to design the public contribution to the research issues.

[24] For example, an achievements page was added that lists four out of five phase 2 games. It will display a scores leaderboard and latest achievements awarded.

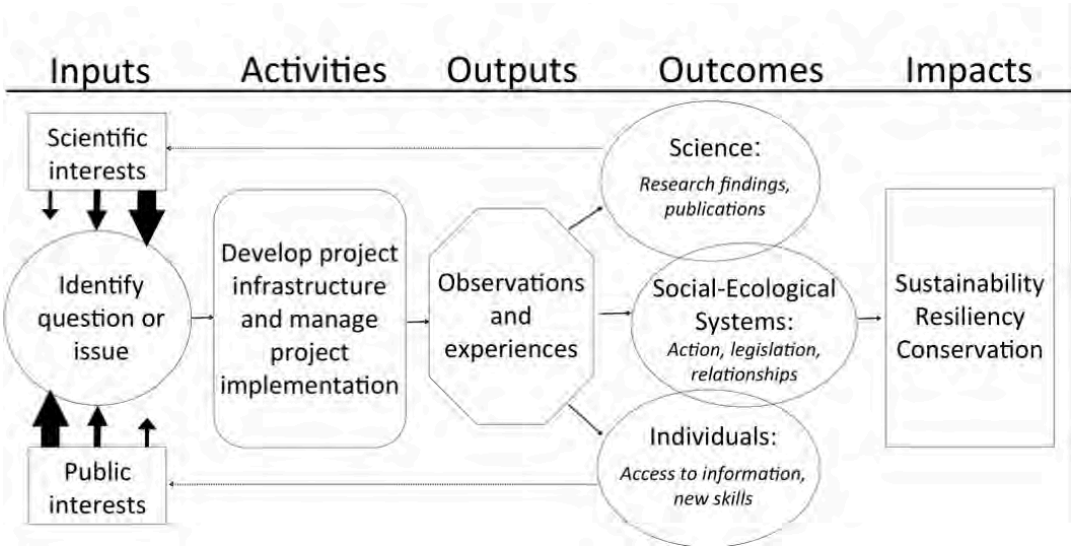

Figure 4.2: Framework of PPSR Projects (Shirk et al., 2012)

Sprinks, Houghton, Bamford, and Morley (2015) went more detailed and discovered that the task workflow design, how tasks are provided to the crowd, is a key element in keeping the scientific interested community engaged for which no best-practice has been developed yet. They found that the task workflow design has to reflect the specifics of the task, the required judgement, the user autonomy, and the coherence of user experience and scientific results.

To leverage the private input, Irwin (2014) proposed that focus has to put on story-telling to foster contributors' engagement. The stories have to address the issues by confronting with both the challenges and the possibilities. For CSFV, this means, the understanding and awareness on software dependency has to be raised as a main goal of the real business goals. However, the discussion has to be held in a non-Western-focused way and needs to cross boundaries (Irwin, 2014; Newman et al., 2012).

To summarize, there is a projected trend of popularity of "average Joe" being part of a scientific community. People want to spend their time and effort productively and contribute in scientific efforts. However, utilizing the "citizen science" trend has to be actively shaped. Experiences from the public, non-profit domain need to be incorporated to the private sector for successful business models. People want to address the needs, but they also want a

deeper understanding about what happens in the background. The "Vision, Mission, Goals" statement of the Citizen Science Association is an indication that citizen science wants to be recognized.[25] They need a platform that facilitates the scientific research, but also recognizes the people's need for recognition. Creators of such a platform need to identify the best practices of already existing projects, and embrace community of interests consisting of experts and volunteers.

4.3 Society and Education

We believe that this shift is not reflecting today's reality of society's skills. In 2008, a study revealed that the United States does not develop the math skills of kids as needed, and that "girls who do succeed in the field are almost all immigrants or the daughters of immigrants from countries where mathematics is more highly valued" (Andreescu, Gallian, Kane, & Mertz, 2008; Rimer, 2008). Another study from the for Economic Cooperation and (OECD) (2013) showed that the math skills of the current young generation in the U.S. in the age between 16 and 24 years rank the last spot with 29 points below young people across all 23 observed countries in math. In the age group 25-34, the U.S. population also only reaches the second last spot.[26] In the Program for International Student Assessment, the so-called PISA ranking, 15-year-old students were assessed in 65 countries. According to Mark DeLoura, former senior adviser for the White House Office of Science, Technology & Policy, the U.S. ranks #36, "just above the average in all categories" (Sparks, 2015).

However, another report from the OECD states that "there are very few countries in the world that are able to make better use of their citizens skills than the United States" (Kis & Field, 2013, p. 3). We assume that business human-assisted formal software verification needs a human-centered approach first to fully take advantage of the player's willingness to voluntarily contribute their time and efforts.

[25]The Citizen Science Association promotes a world where people understand, value, and participate in science. Several high-ranking non-profit organizations have already joined the association (Citizen Science Association, 2014).

[26]Rustad (2011) confirmed this observation on a far smaller scale for Monterey County only. The study found that 28% of the county's adult lack literacy skills (Panetta Institute, 2015).

4.4 Conclusions on Trends

Gamification and crowdsourcing in form of citizen scientists offer lucrative opportunities. Both elements need sophisticated application that needs nurture and continuous adaption. The players have to gain center stage. Currently the focus of the crowdsourced formal verification approach lies on the technology. We believe that maturity, as we defined it in the beginning, cannot be achieved. Success stories evolve from the interaction between the innovative ideas and the people.

The CSFV project is shifting its focus to the citizen scientists. This workforce may have the skills necessary to easily contribute to the current games. But, as the lessons learned from phase 1 showed and the studies about the state of education showed, the overall society is most likely not able to voluntarily contribute in numbers that are needed to produce the amounts of user assertions necessary to proof even small code.[27]

Based on the insights of Chapter 3 and Chapter 4, we recommend to shift the focus to the people's/ society's behavior. A human-centered approach allows to see the bigger picture and may unbound the limits of a too science and technology focused community.

[27] The BIND code has only 300 KLOC.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5:
A Human-Centered Business Model for Crowdsourced Formal Verification

> The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency. (Bill Gates)
>
> —Bill Gates, Philanthropist

> The computer is incredibly fast, accurate, and stupid. Man is unbelievably slow, inaccurate, and brilliant. The marriage of the two is a force beyond calculation. (Leo Cherne)
>
> —Leo M. Cherne, Economist, 1977

An innovative idea cannot traverse along the Hype Cycle by only maturing technically and though academic discussion. Public visibility and maturity, the "consumer acceptance of the basic service idea, by widespread believe that the products of [] will perform satisfactorily, and by enough familiarity and sophistication" (Paliwoda & Thomas, 1998) needs to be promoted by real performance stories. A human-centered business model strives for solutions that are desirable, feasible, and viable.

According to Osterwalder and Pigneur (2010), a business model is defined as "the rationale of how an organization creates, delivers, and captures value." We understand that the value in crowdsourced formal verification is not the monetary portion of the formal software verification business. The value of the crowdsourced formal verification approach is to leverage the contribution of non-experts to spend their time and effort in producing user assertions that contribute to the continuously improving automated verification technology,

which finally will result in lower costs for software verification.[28] Unintentionally, players share not only potential solutions to the goals that e.g., FRAMA-C produced, but also share insights about humans think in terms of identifying patterns and producing equations. Business can use these insights to further develop the technology behind the formal verification tools.

5.1 Defining a Human-Centered Approach

In the introduction, we argued that we would follow McLuhan's understanding of the relationship between technology and humans. We define humans in accordance with Kling and Leigh Star (1998) as "individuals and their cognitions [that] include the activity and interactions of people with various groups, organizations, and segments of larger communities." This definition is valuable to our understanding as it allows to focus on people's behavior as a driver of (inter-)action while contributing to the formal verification of software. The Field Guide to Human-centered Design by Ideo.org states that "[h]uman-centered design offers problem solvers of any stripe a chance to design with communities, to deeply understand the people theyre looking to serve, to dream up scores of ideas, and to create innovative new solutions rooted in peoples actual needs" (Design Kit, 2015). This definition ties in our actual debate about effectiveness of crowdsourcing and gamifying the formal verification of software. Differently to an organization-based approach, like that of SAP Enterprise Ressource Planning software, a human-centered approach would consider the users preferences, needs, and desires.

5.2 A Primer For an Human-Centered Approach

The Verigames.com platform offered interested people during phase 1 an opportunity to pursue a new way of contributing to formal verification of software. Based on the available data for one of the five games, we showed that the contribution of these people to the process was limited by the formal verification back-end and the front-end, and the actual skills of players. This findings should not be understood as negative critique to an experiment that explores the crowd-assisted formal software verification for the first time. Nor is it meant to be a critique on the way, the Verigames.com platform offers these CSFV games. The five games have been the first one of their specific genre that merged formal methods

[28] Because it reduces dependency on human experts.

and tools, gameplay, gamification means, crowdsourcing technology, and voluntary willingness in a way that had never been done before. We found that the crowd was able to fully verify 25.08% of the available tasks, and that a lot of free effort could not be converted toward efficient formal verification. We also found that the games still do not attract many players, like other games of the serious games category do. We also looked at the trends of some of the elements that contribute to the CSFV right now and their future prospects. We thereby came to the conclusion that these issues may be resolved within the next 2—5 years due to technological advances, and the impact of more effective use of gamification and crowdsourcing, or impacts of other, not yet considered developments that may reach the plateau of productivity at the same time.

However, we believe that one issue cannot be resolved by just adding technological improvements. The key resource that the crowdsourced formal software verification approach depends on, is not the technology. Also there is evidences that the CSFV players may belong to specific user groups (cf. Tellioglu, 2014; DARPA, n.d.), it is the non-expert that contribute voluntarily for different reasons to the process. It has not been proven so far, whether the CSFV players are more likely game players, citizen scientists, or people interested in mathematics. We recommend to shift the focus on this element of the equation, as it may be the most difficult element to improve. Shifting the perspective towards a human-centered approach may in the medium-run be a game changer for CSFV. Games that train or help to improve math skills can make the difference. The United States government (USG) already put education by video games on its task list. On the 4th White House Science Fair 2014, a national STEM video game challenge was included. DeLoura justified the President's interest in the "promise of games [] the potential outcomes of games (Sparks, 2015).

5.3 Example of Successful Human-Centered Crowdsourced Projects

The Norwegian company WeWantToKnow developed three puzzle games, which also fall under the serious games category. Since 2012, the games caught worldwide recognition, because they enabled children to secretly learn algebra and geometry. Game-based learning is a traditional and well-tested approach to deep and effective learning (Gee, 2013). The

games were designed by Jean-Baptiste Huynh, a math teacher, and Dr. Patrick Marchal, a cognitive scientist . The Dragonbox games utilize the fact that information learning can better be processed by students by using these information in problem-solving. Similar to some of the CSFV games, the Dragonbox games also use symbolic reasoning to solve the algebraic problems. In accordance with the Goal-Setting Theory, Gee (2003) argued that not advanced 3D graphics would catch the attention of an interested player, but its underlying architecture that challenges " the outer and growing edge of a player's competence" and stays doable.

The Dragonbox games are subject to ongoing research and testing by the Center of Game Science from the University of Washington (WeWantToKnow, 2015). The company and the university also founded the Massive Interactive Learning Events program, where students from all over the world come together to challenge each other in groups.

Another example from the education sector is Duolingo. Duolingo is a crowdsourced language-translator that offers free language-education to customers worldwide. What started as a spin-off from the Carnegie-Mellon University has today 30 million customers. The founder Luis von Ahn reported in January 2014 a retention of about 8.5 million active users out of 20 million. Companies pay to get documents translated, while a free labor force interested in language learning translates these documents. While Duolingo advertises to offers the language training free forever, the company also reaches out for other profitable markets like language certification (Duolingo, n.d.; Liu, 2015; Olson, 2014; Root, 2014).

Both games games are interesting examples for a business model. Formal Verification games need player that have math skill, and are interested in applying math. The Dragonbox games provide these skills and prepare a sustainable customer base, by focusing on the needs and behavior of their players. Additionally, they convince children's parents of the desirability of games with a purpose. Duolingo provides insights to a successful crowdsourced business model. Both games cannot considered "best practice," but they provide reasonable orientation how crowdsourcing and gamification can encourage a large customer base.

5.4 Proposal of a Human-Centered Business Model

We propose a human-centered business model. It may fix the current lack of user numbers. We apply the business canvas methodology that allows us to draw a more strategic, holistic picture of the elements that need to be considered for a sustainable business. Referring to the Framework of PPSR by (Shirk et al., 2012), we describe a potential solution to the CSFV games based on a two-sided market concept that brings together the financial power of the software-dependent customers with the crowd-assisted power of people interested in education. This proposal can be used by the Verigames.com platform, but is also generic enough that it can be used by any business offering crowdsourced formal verification of software. [29,30]

5.4.1 Customer Segments

Customer segments define the different groups and organizations a company aims to reach and serve. CSFV games serve two independent markets. For the purpose of this thesis we divide the customer segments by their needs. One segment consists of businesses and organizations that are dependent on security and safety critical software. Another segment is the player segment.

The player segment should not only address playing voluntarily free-games on the Internet, but exploit the growing demand on the educational sector. Games that train math skills currently may fit multiple purposes. First, formal methods can be more easily applied to the game story and game design (cf. DARPA, n.d.). Second, developing games that teach math skills opens a solvent new customer base, as parents are more likely to pay for serious games that educate their children. Parents spend over $2.7 billion per year on children apps (Ante, Troianovski, & Vascellaro, 2012). Additionally, the games add a customer base that spends lots of time in games and will most likely sticks to the familiar behaviors over the years. The average child spends three hours per day with games (ZeroDesktop Inc., 2014). Also, it prepares the future generation to develop the skills necessary in a software dependent world.

[29] In this section, we do not refer to the DARPA Verigames, but we refer to verification games in general to distinguish the goals of the DARPA experiment and the intentions of this business model.

[30] The described business plan elements focus mainly on the on the education business, rather than the verification business.

5.4.2 Value Proposition

The value proposition describes how a business model tries to solve customer needs. Tying verification games with an educational incentive provides an immediate value to the players. The keywords for the value proposition of our business model for the players are motivation and convenience, good education in an inexpensive way.

Education user are motivated by learning, part intrinsic, part extrinsic. Educational crowdsourced formal verification games provide intrinsic motivation. Ryan and Deci (2000) summarized three needs that drive the intrinsic motivation: competence, relatedness, and autonomy. Therefore, such games have to embrace the self-determination of the customers. Games, per se, can satisfy these needs, and even rise their level if they address the relevant regulatory processes. According to Ryan and Deci (2000), these are interest, enjoyment, and inherent satisfaction. Educational games need to address these regulatory processes. Additionally, Kaplan (2010) argues that "when curiosity, independence, and exploration result with experiences of mastery and meet the approval and encouragement of parents or teachers, children experience pleasure, feel competent and in control of their environment, and have stronger intrinsic motivation for the domain or activity" which conclusively may lead to improved retention rates.

Educational games offer convenience to parents. A study from the fourth quarter 2011 showed that in 2011 seven out of 10 children use a tablet computer (Nielsen, 2012). This was a 9% increase from the third quarter in the same year. A poll in 2013 by CouponCodes4u.com revealed a similar result. About 58% of the U.S. parents tend to babysit their children through the use of technology (Taylor, 2013).[31] The tendency of modern parenting opens a new customer base, looking for convenience for their lifestyle which also allows to justify it by easy accessible education for their children.

Educational crowdsourced formal verification games that implement sophisticated gamification provide extrinsic motivation. Although some academics argue that extrinsic motivation has detrimental effect (Kohn, 1999), others argue that rewards even contribute to the intrinsic motivation (Eisenberger & Cameron, 1996).

We also propose a reduction of dependency on the volunteering workforce of citizen scien-

[31] The poll included more than 2,400 parents across the country.

tists. We showed in Chapter 4 that citizen scientists are likely to develop more sophisticated aspirations, aspirations that may be more difficult to satisfy by games. Scientific citizens are also motivated by their involvement and their contribution based on altruism. Research on the motivation for participation in technology-mediated social participation and for the practice of citizen science by Nov, Arazy, and Anderson (2014) revealed that collective motives, and reputation drive the quality and quantity of users' contribution (cf. also Curtis, 2015; Wasko & Faraj, 2005). However, the motivations between scientific citizens and crowds interested in education are different. As we showed above, education is triggered by intrinsic motives. Nov et al. (2014) showed that intrinsic motivation additionally determines the quantity of users' contribution. We conclude, that education motivation is a better motivational basis for achieving higher quantitative goals for the crowdsourced formal verification, as it is mainly driven by the intrinsic, self-determined needs of humans. The education-driven motivation provides a stronger basis for ramification and further engagement than the scientific citizen approach. Shifting the value proposition towards education opens the door to a customer base - parents and their kids - with intrinsic motivation, which is why the intuition about setting these games up as educational games has attraction to achieve higher quantitative contribution.

5.4.3 Key Resources

Key resources may be physical, intellectual, people, or financial assets. Our business model focuses on the intellectual and human resources necessary to create the value proposition that we want to offer. One aspect of the proposed value proposition of an human-based business model is the aspect that we care about how seamless the product will integrate into our lifestyles and how well a different way of crowdsourced formal software verification can be applied.

Developing games for formal software verification that do not target citizen scientists, but people interested in education (parents, teacher, and kids) requires professionals who are able to translate the market needs, and scientific aspects merged in this complex product. For example, the business needs experts in formal methods, like computer scientists and mathematicians to generate a successful formal verification. Also, the business needs to include people who understand psychological behavior and learning methods of humans. Third, the business needs game developers who are able to generate attractive game play

for the end-user that integrates the environment of the players (such as parents of the kids, teacher in the classroom) (Haklay, 2013).

Additionally the business needs to establish a forum (e.g., Internet platform) that embraces conversion of interest into revenue, game play, and community exchange.

5.4.4 Key Activities

Key activities describe the core of a business' actions. Based on our value propositions, we have to provide a complex set of business activities in order to bring together two different customers with different requirements.

First, problem solving capabilities have to be offered to satisfy customers looking for formal software verification. Customers may need additional consultancy in defining the specifications of their code. Customers may also have different needs for security and protection of their code, which requires different approaches to secure their interests.

The second key activity is the development of education software, to exploit the motivation of a customer base looking for education. Formal methods for software verification have to be translated in educational games that apply to different user age group. The software development must be supported by strong academic proven insights about human learning.

Third, platform and network activities have to be established to link the two different customer segments, and to provide a continuous discussion about enhancement in e-learning activities. We must provide a platform that allows to connect the different interests, bringing together formal verification competence, game development expertise, and experts from the educational sector.

5.4.5 Key Partnerships

Key partnerships summarize the network of suppliers and partners that make the business work. Partnerships have to be established to reduce risk, and to acquire resources that are provided less expensively on the market.

A business might not own all resources and capabilities to provide video game competence, formal verification competence, and platform competence. A potential spin-off should incorporate these capabilities by partnering with specialists in the respective fields. Keeping

the core competence of building the vision for the educational verification games should be the main focus, while including the above mentioned competences through partnerships enables to optimize economies of scales.

We recommend building partnerships with educational institutions to study long-term impact of educational games embracing mathematics. Also, partnerships to school districts help to establish a testbed that allows feedback of new features before exposing the games to the market.

5.4.6 Customer Relationships

To influence customer acquisition and retention, we need to define the type of relationship for each customer segment. In Chapter 3, we confirmed the results of a former study (cf. Tellioglu, 2014) that customer acquisition and customer retention were weak points of Verigames. An effectiveness study about Duolingo from 2012 found that many of their participants dropped out of the study or spent less than two hours studying Spanish (c.f. Vesselinov & Grego, 2012). The study recommended that Duolingo should implement some kind of advisable means that would encourage the players to keep studying. Duolingo implemented daily push messages telling the players how to reach the study goals. These messages are personalized and limited for a specific period of time to not annoy the users. This is a good example of sophisticated personalized automated services. In Chapter 4, we identified that also citizen scientists are looking for more sophisticated recognition of their word. The customer relationship needs to reflect that sophistication in terms of services.

Another tool is the creation of communities. The DARPA CSFV project increased the social media engagement on Verigames.com for phase 2. Communities enable companies to be more involved with their customers by allowing the customers to interact directly with each other. An example is the Garmin User community. [32] Users discuss their current success stories, share their expectations and worries, and mention concerns and problems. The Garmin company participates on a small scale in the discussions, but listens continuously to better understand the customer base.

Communities can be used to create and gather ideas for business development. Duolingo uses the community tool for their Duolingo Incubator. Bilingual people from all around the

[32] The Garmin Forum can be found under https://forums.garmin.com/forum.php .

world get connected to create new language courses.[33] This way, Duolingo not only serves the current customers with new courses, but also create two new key resources: ideas and free labor force.

From a human-centered perspective, customer relationship is a crucial key element that deserves strategically carried out planning and operations. Shute, Rieber, and Van Eck (2011) argues that "[t]he ability to work creatively and effectively with others toward a common goal is an important 21st century skill that is emphasized in good games." The above "best practices" show that a thorough customer relationship handling sets the foundation of a sustainable business.

5.4.7 Revenue Streams

The revenue streams are determined by the market values of the education sector, educational games industry revenues, and the revenue models chosen for games. The most common revenue models we found on the market for serious games are purchase or pay-per-use, and freemium models.[34] The Dragonbox games represent the one-time-fee fixed pricing model, while Duolingo represents the Freemium model.

The market for eduction in general and education games in specific is rising worldwide. According to IBIS Capital, a London-based investment bank, the global market for education was $4.4 trillion in 2013 (Global education market, 2015). IBIS Capital forecasted a growth by 23% by 2017. The Global Industry Analysts Inc. forecasted the global market for e-learning to reach $107 billion by 2015 (PRWeb, 2012). Ambient Insights found that the market for educational games will rise from $1.5 billion in 2012 towards $2.3 billion by 2017 (Takahashi, 2013). Takahashi (2013) calculated that, based on a five year compound annual growth rate of approximately 9.2%, the "self-paced eLearning market should see estimated revenues of $49.9 billion in 2015" (Takahashi, 2013; Ambient Insight, 2013).

The website Gamesandlearning.com offers additional insight in the current prospective market dynamics of educational games (cf. Games and Learning, 2015). Besides parents, education professionals are increasingly interested in applying video games in their

[33] The name Incubator also implies the Duolingo quality management approach. The courses run through three distinct phases until they are considered to meet the defined quality criteria.

[34] Freemium describes a free-to-play revenue model that tries to convert users from free playing in playing on basis of subscription or usage fees.

portfolio. A study by the Stanford Graduate School of Education found that limited video gameplay shows significant math improvement (Pope, Boaler, & Mangram, 2015). The study was not based on Dragonbox, but on a less well-known the game called Wuzzit Trouble.

Both examples used for best-practices in this thesis, Dragonbox and Duolingo, try to spread their revenue base through different channels. We can assume that this will have influence on the revenue stream in the mid-term. Wilson argues that Duolingo will additionally charge for their services as soon as they are able to place their products, like the Duolingo Test Center, at schools and employers (Wilson, 2014; Straumsheim, 2014).

Similar to the Duolingo approach, a CSFV approach should establish a two-sided market where educational users get free access to the games, while the formal verification customers pay for the games development.

A CSFV approach should further exploit the freemium model to get the most users. Todays society is used to and expects to have free access to products.

If you can incorporate competitive elements and social comparisons them you might also get a network effect which might speed-up and increase adoption. This is the classic risky bet for the game producer (eg Darpa, at this point) because you have to invest in making the education game pretty good, and invest in getting it to diffuse, and then only much later do you reap the potential benefits of a large-enough user base to be useful for verification purposes.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 6:
Conclusion

> Ten or twenty years from now we'll likely have a more universal theory of which tests to write, which tests not to write, and how to tell the difference. In the meantime, experimentation seems in order.
>
> —Ken Beck, Software Engineer, 2008

This thesis examined the maturity of crowdsourced formal verification games played on the Internet platform Verigames.com during December 2013 and May 2015. The data set based on raw data collected from the back end of the game Stormbound and on results reported from the different game teams.

With the crowdsourced formal verification games, DARPA explored a completely new domain of semi-automatic formal verification techniques. DARPA does not expect their programs to be always successful. Neither is a transition to the "real business" a pre-defined outcome (Hanisch, 2010). We found that the phase 1 games still showed some technical deficiencies that make one dubious about the maturity of the approach so far. Also the available data set is limited. The evaluation of the data revealed that the games suffer from limitations on the back end side, which might be resolved in the future. Currently, more research is needed to result in technical maturity of the used program analyzers for C programs (e.g., FRAMA-C), and the combination of crowd-assisted assertion generation and automatic proving. Also, the current way of assertion collection does not allow a substantial conclusion about the code being free of certain bugs, as long as assertions from the play of games are collected independently from each other.

We also found that the front ends, the actual games, are already fairly mature from a technical stand-point, but that the phase 1 games might have been too much of a niche product to convince a casual user group with general math skills. Stormbound tried to avoid confronting users with mathematics, and, therefore, suffered from reduced verification success. Other games, like Xylem, were locked in the "math-intense" user niche. Both approaches

converted to a even more mathematical exposed user experience in the games produced for phase 2. We also found that the rudimentary use of gamification is a key element in embracing user's attention and to bind their loyalty.

We also looked at the future prospects of gamification, crowdsourcing, and the status quo of education. We showed that the above mentioned technologies will probably reach the plateau of productivity in five to 10 years. We expect that these technologies will shift the direction of a more sophisticated user experience in formal verification.

However, we are convinced that a focus on a science-oriented crowd does not provide the necessary user counts to gain enough assertions to fully verify excessive amounts of code in the long run. A successful crowd-assisted approach in formal software verification has to consider the reality of society. Society is not defined by a niche, but by the general trends in education and knowledge. As a consequence, we developed a human-centered business model that takes the reality of the current society into account.

We recommend for further development of the games to primarily focus on players with an interest in education. The lessons-identified during phase 1 teach us that if we want to exploit the human resource subliminally for business needs, we need to put people's behavior in the focus of our research. We recommend early adaptors to target players' needs, to embrace their social bent by fully utilizing their interest in learning and cheap, convenient education by means of gamification. We assume that this will unbound the dependency on having to pay them, e.g., on Amazon Turk, as satisfied players will spread the word resulting in a growing and even substantial count of users.

The recommended business model tries to utilize a two-sided market. The main revenue has to result from businesses who want their software to be formally verified. This allows to offer the educational games as a freemium model, similar to Duolingo, though convenience and education is still "worth" something to the current parents generation.

Either way, it is too early to conclude if a crowdsourced formal verification approach might result in lower costs of formal verification or less dependency on the scarce resource of human experts. The technologies enabling a more performant semiautomatic and crowd-sourced formal verification of software is still in its infancies. We conclude that it will take

another two to five years until DARPA's gamified approach of formal verification of software will be mature enough to justify a financial investment, and that the normal human, not formally trained in verification methodology, together with computers can make our software-dependent world safer.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A:
Figures of the Quantitative Analysis

A.1 Figures on Players Contribution

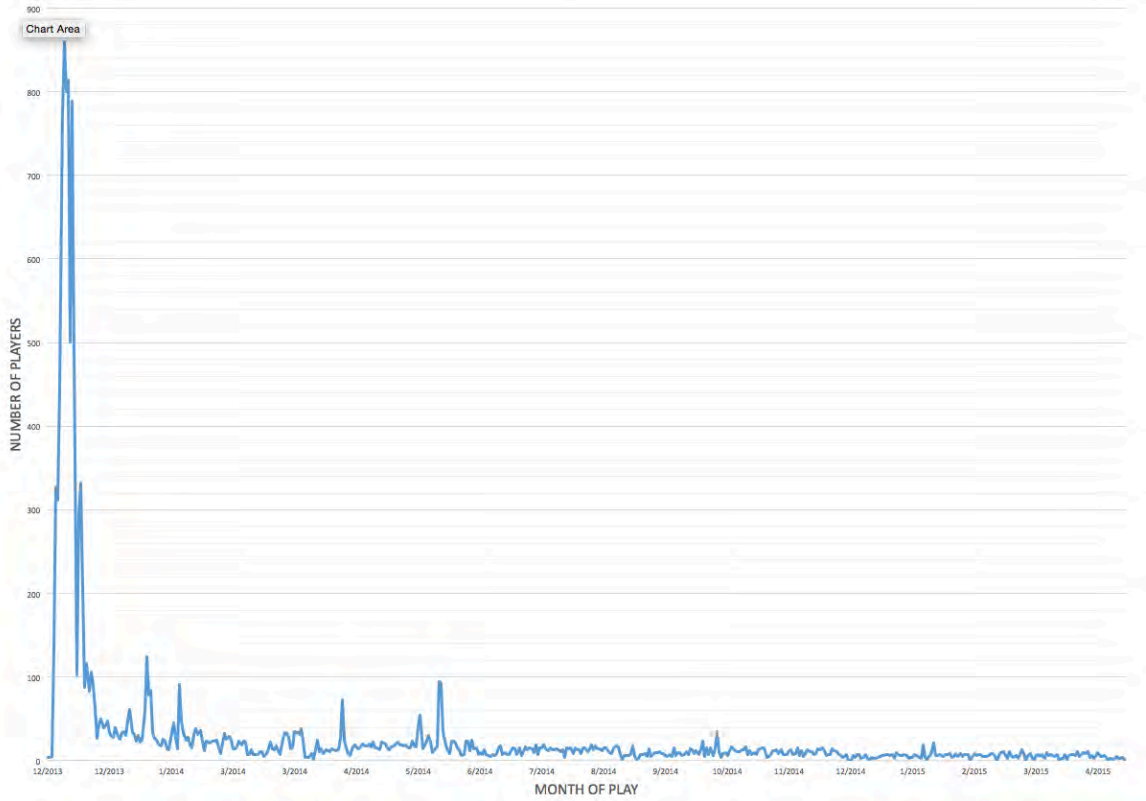

Figure A.1: Daily Average Users During Phase 1

A.2 Figures on Quantitative Analysis of Games Data

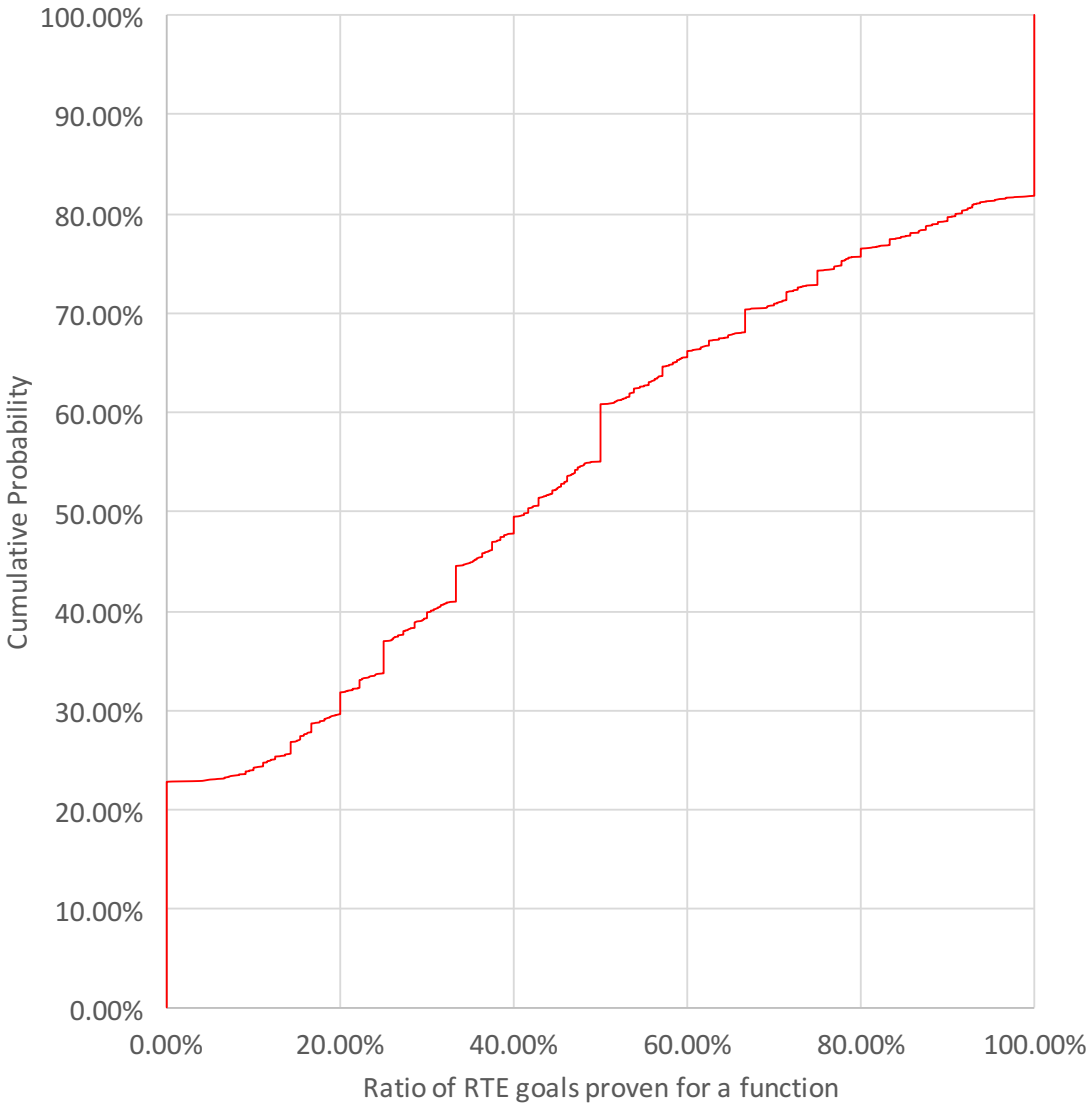

Figure A.2: CDF for Results of proofed RTE Goals

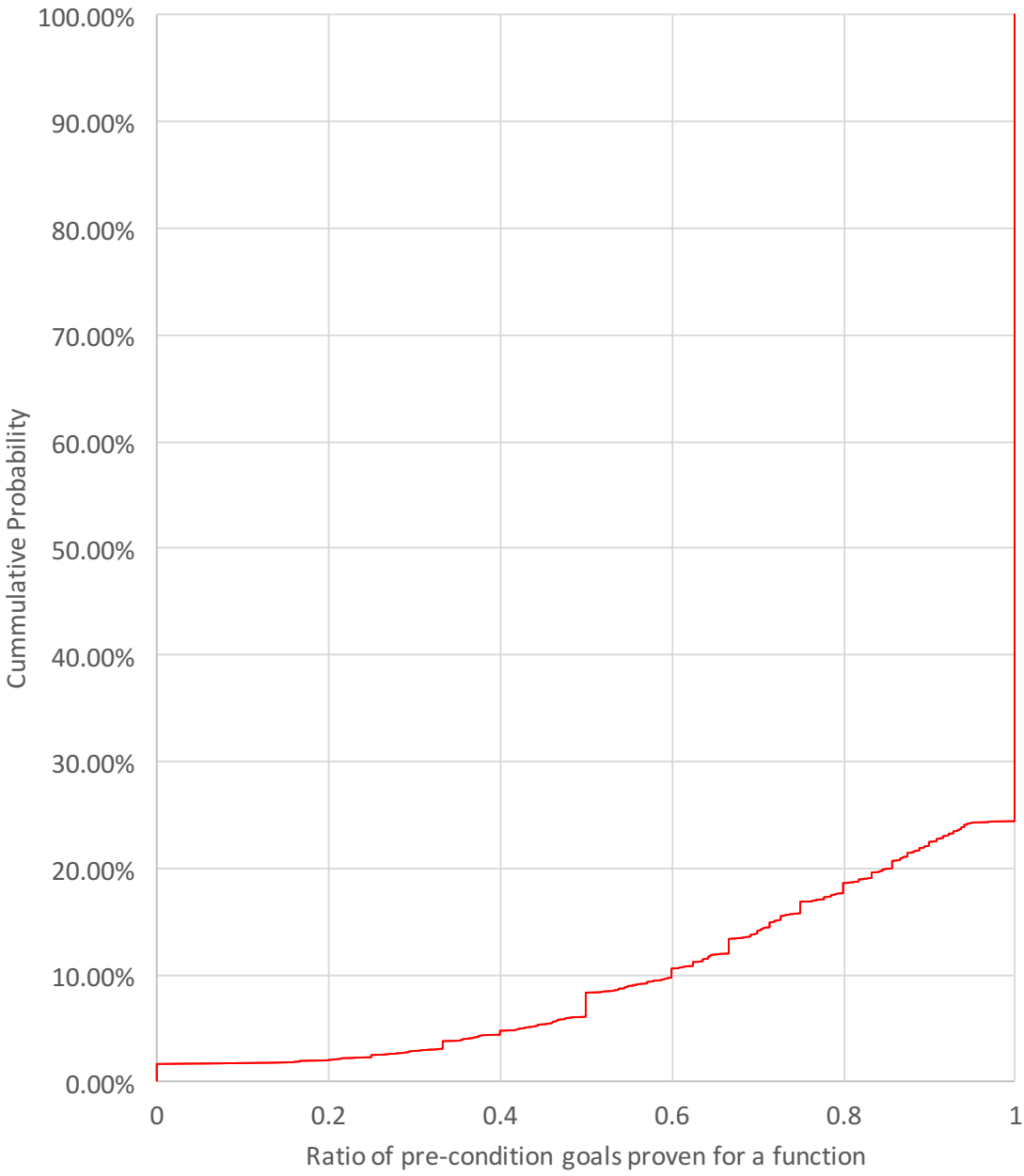

Figure A.3: CDF for Results of Proofed Pre-Condition Goals

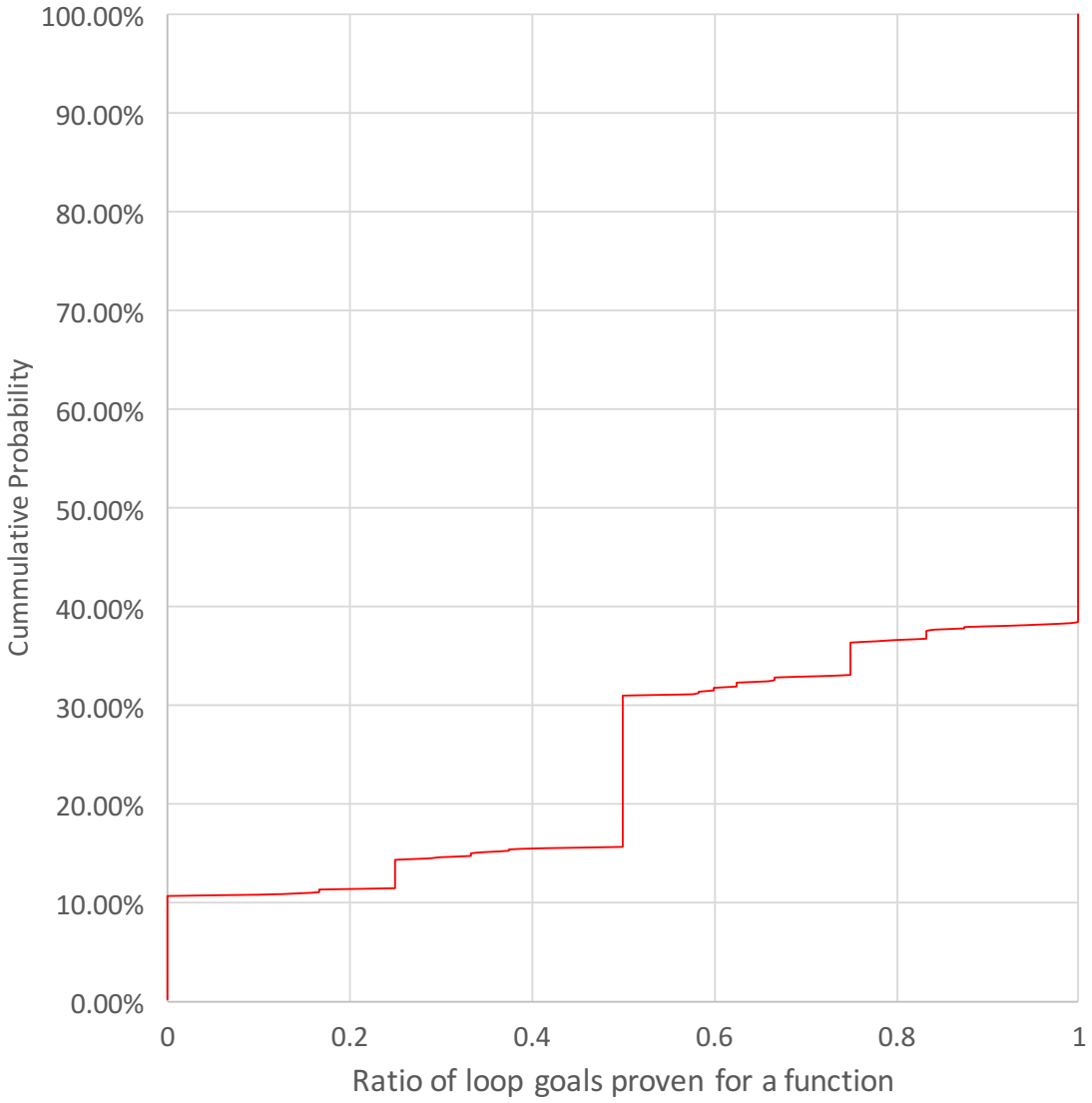

Figure A.4: CDF for Results of Proofed Loop Invariant Goals

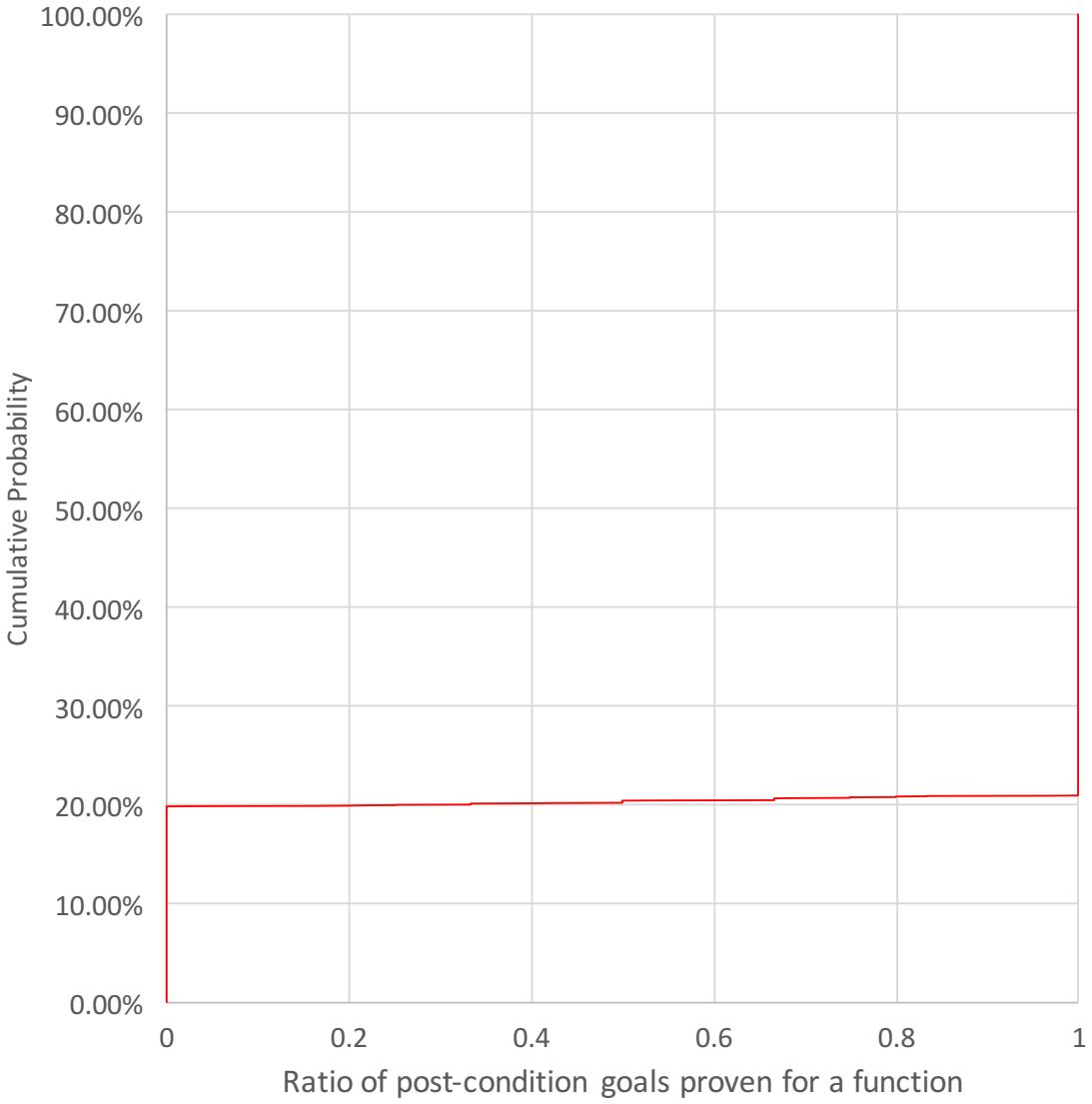

Figure A.5: CDF for Results of Proofed Post-Condition Goals

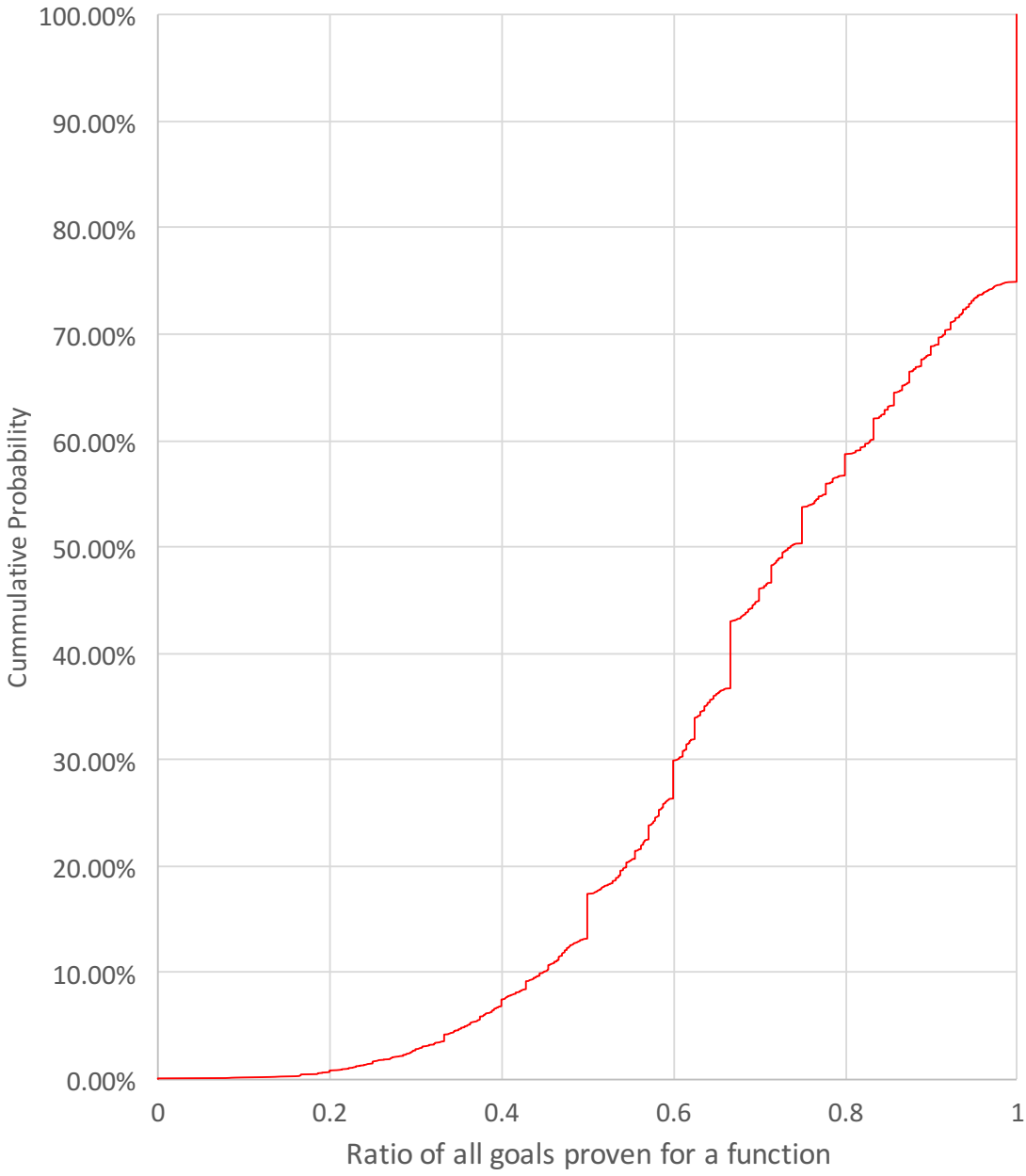

Figure A.6: CDF for Consolidated Results of Proofed Goals

A.3 Figures on Stormbound's Engagement Rate

The distribution in Figure A.7 in the Appendix shows a dominating low ER on most days over the phase. Figure A.8 shows the daily average ER, while Figure A.9 shows the monthly average ER and participation during phase 1.

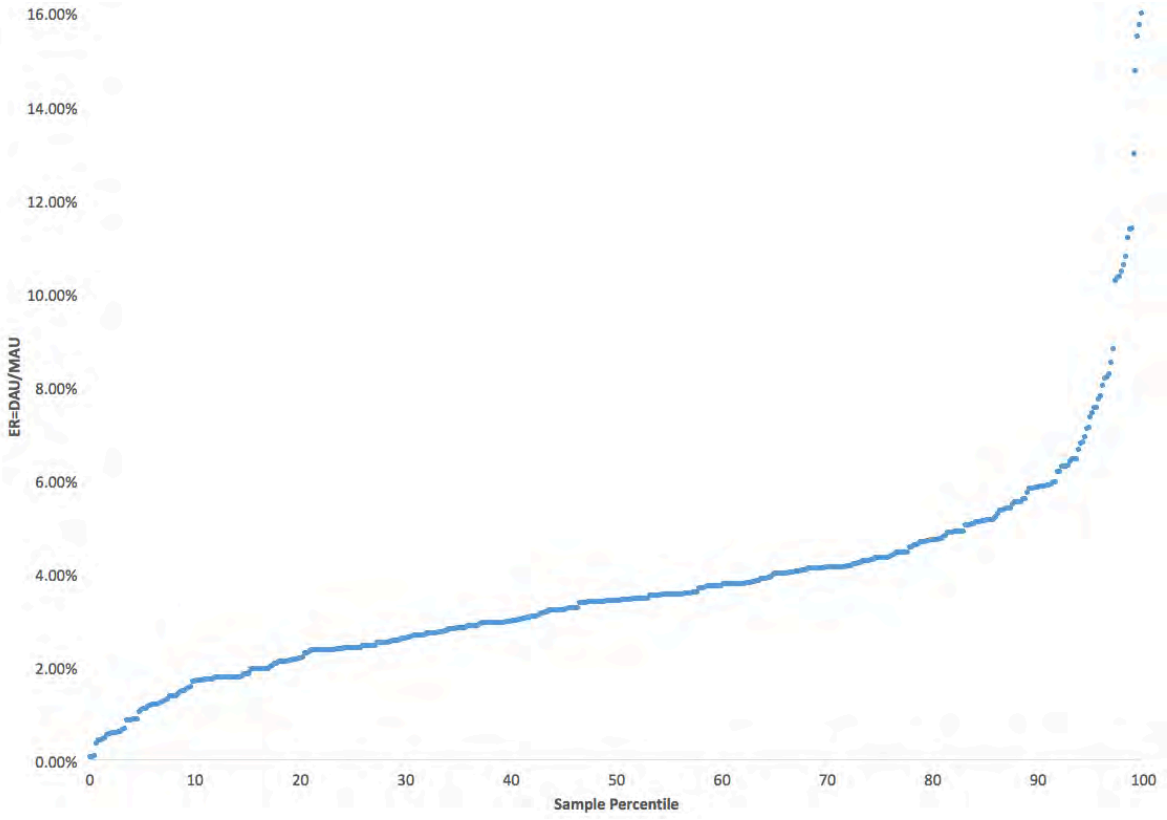

Figure A.7: Distribution of ER from December 2014 to April 2015

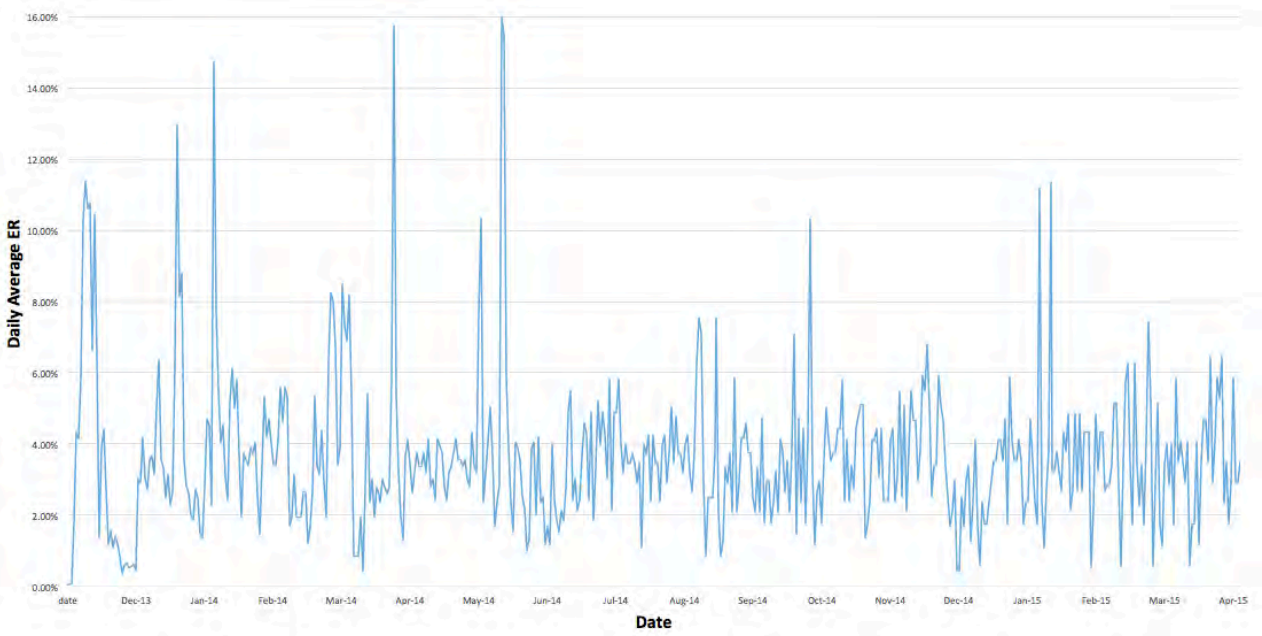

Figure A.8: Daily Average ER During Phase 1

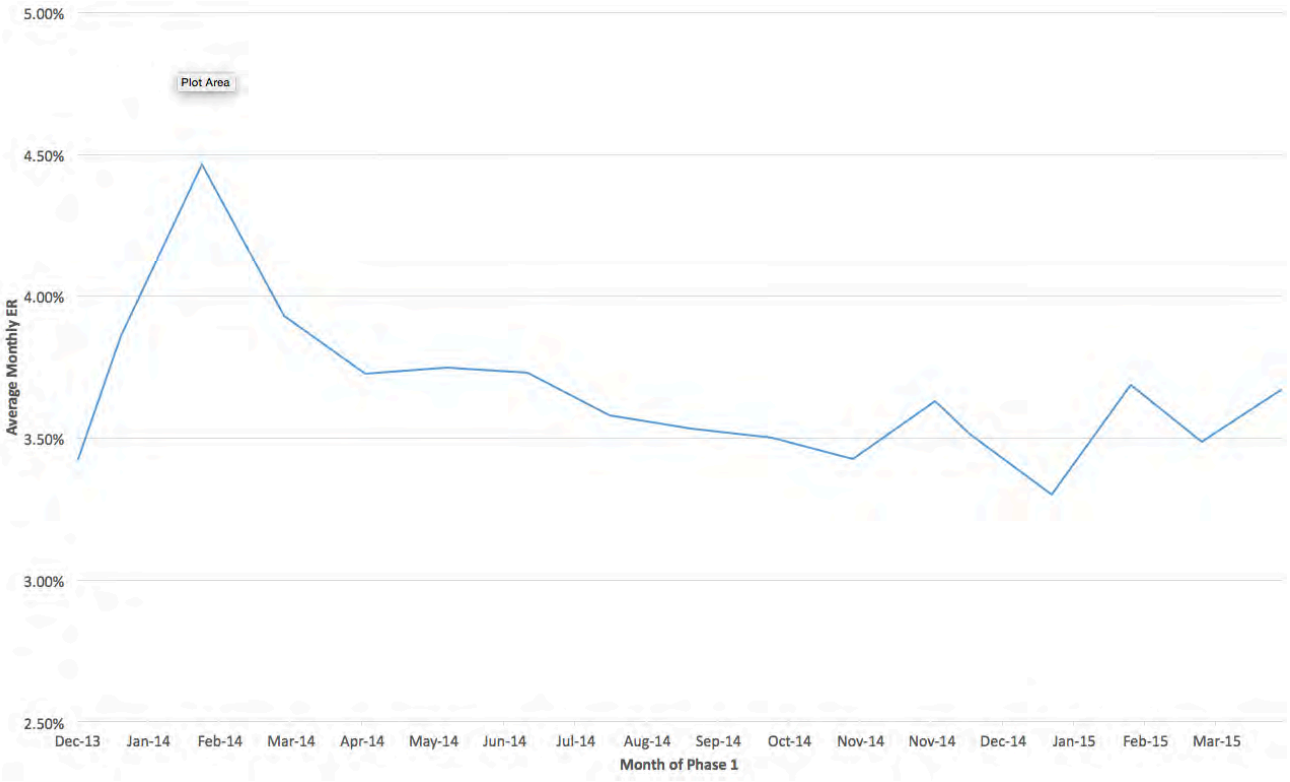

Figure A.9: Monthly Average ER and Participation During Phase 1

THIS PAGE INTENTIONALLY LEFT BLANK

References

Ambient Insight. (2013). The 20122017 worldwide mobile learning market. Retrieved from http://www.ambientinsight.com/resources/documents/Ambient-insight-2012-2017-worldwide-mobile-learning-market-executive-overview.pdf

Anderson, J., & Rainie, L. (2012). The future of gamification. *Pew Research Center's Internet & American Life Project*. Retrieved from http://www.pewinternet.org/2012/05/18/the-future-of-gamification/

Andreescu, T., Gallian, J. A., Kane, J. M., & Mertz, J. E. (2008). Cross-cultural analysis of students with exceptional talent in mathematical problem solving. *Notices of the AMS*, *55*(10), 1248–1260.

Ante, S. E., Troianovski, A., & Vascellaro, J. E. (2012). Mom, please feed my apps! *Wall Street Journal*. Retrieved from http://www.wsj.com/articles/SB10001424052702303753904577452341745766920

Barsky, J. (2012). Gamification Engage Your Customers with Fun and Games [Blog]. *Hospitality Net*. Retrieved from http://www.hospitalitynet.org/news/4055106.html

BBC News. (2014). Nasa to hack mars rover opportunity to fix amnesia fault. *BBC News*. Retrieved from http://www.bbc.com/news/technology-30642548

Beck, K. (2008). How deep are your unit tests? Retrieved from http://stackoverflow.com/questions/153234/how-deep-are-your-unit-tests

Blaney, E. (2014). Five intrinsic motivators and how they impact employee engagement. *Bunchball*. Retrieved from http://www.bunchball.com/blog/post/1591/five-intrinsic-motivators-and-how-they-impact-employee-engagement

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York, NY: WW Norton & Company.

Burke, B. (2013). The gamification of business. *Forbes*. Retrieved from http://www.forbes.com/sites/gartnergroup/2013/01/21/the-gamification-of-business/

Burton, B., & Willis, D. A. (2014). Gartners Hype Cycle special report for 2014. Retrieved from https://www.gartner.com/doc/2816917/gartners-hype-cycle-special-report

Butler, R. W. (2001). What is formal methods? Retrieved from http://shemesh.larc.nasa.gov/fm/fm-what.html

Chitale, R. (2009, October 13). Doctors Shocked by radiation overexposure at Cedars-Sinai. Retrieved from http://abcnews.go.com/Health/CancerPreventionAndTreatment/doctors-shocked-radiation-exposure/story?id=8818377

Cialdini, R. B., Goldstein, N. J., & Martin, S. J. (2008). The science of getting a "yes." *NPR.org*. Retrieved from http://www.npr.org/templates/story/story.php?storyId=93872977

Citizen. (2015). In *Oxford English Dictionary*. Retrieved from http://www.oed.com/view/Entry/ 33513?redirectedFrom=citizen +science#eid316619123

Citizen Science Association. (2014). Vision, mission, goals. Retrieved from http://citizenscienceassociation.org/overview/goals/

Clarke, E. M., & Wing, J. M. (1996, December). Formal methods: State of the art and future directions. *ACM Computing Surveys*, *28*(4), 626–643. Retrieved from http://doi.acm.org/10.1145/242223.242257

Collins, M. (1998). Formal methods. Retrieved from http://users.ece.cmu.edu/~koopman/des_s99/formal_methods/

Cook, D. (2006). Lost garden: What are game mechanics? *Lostgarden.com*. Retrieved from http://www.lostgarden.com/2006/10/what-are-game-mechanics.html

Culkin, J. M. (1967). A schoolman's guide to Marshall McLuhan. *Saturday Review*, *50*, 20–26.

Cuoq, P., Kirchner, F., Kosmatov, N., Prevosto, V., Signoles, J., & Yakobowski, B. (2012). Frama-C. In G. Eleftherakis, M. Hinchey, & M. Holcombe (Eds.), *Software engineering and formal methods* (Vol. 7504, p. 233247). Berlin, Germany: Springer. Retrieved from http://dx.doi.org/10.1007/978-3-642-33826-7_16 doi: 10.1007/978-3-642-33826-7_16

Curtis, V. (2015). *Online citizen science projects: An exploration of motivation, contribution and participation* (Doctoral dissertation, The Open University). Retrieved from http://oro.open.ac.uk/42239/

Daikon dynamic invariant detector. (2015). Retrieved from http://plse.cs.washington.edu/daikon/

Dean, D. (2011). Crowdsourced formal verification. Presented at the CSFV Proposers Day briefing, Washington, DC.

Defense Advanced Research Projects Agency (DARPA). (n.d.). Lessons learned in game development for crowdsourced formal verification. Retrieved from https://basecamp.com/1901825/projects/672654/messages/40069089

Defense Advanced Research Projects Agency (DARPA). (2013). Crowd sourced formal verification (csfv). Retrieved from http://www.darpa.mil/Our_Work/I2O/Programs/Crowd_Sourced_Formal_Verification_(CSFV).aspx

Design Kit. (2015). Design kit. Retrieved from http://www.designkit.org/resources/1/

Deterding, S., Khaled, R., Nacke, L. E., & Dixon, D. (2011). Gamification: Toward a definition. In *Chi 2011 gamification workshop proceedings.* Retrieved from http://gamification-research.org/wp-content/uploads/2011/04/02-Deterding-Khaled-Nacke-Dixon.pdf

Dorling, A., & McCaffery, F. (2012). The gamification of spice. In *Software process improvement and capability determination* (pp. 295–301). Berlin, Germany: Springer.

Duolingo. (n.d.). Learn Spanish, French, German, Portuguese, Italian, and English for free. Retrieved from https://www.duolingo.com

Easterbrook, S. (2010). The difference between verification and validation [Blog]. *Serendipity*. Retrieved from http://www.easterbrook.ca/steve/2010/11/the-difference-between-verification-and-validation/

Eisenberger, R., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American psychologist*, *51*(11), 1153.

Enterprise Gamification Consultancy. (2015). Facts & figures [Wiki]. Retrieved from http://www.enterprise-gamification.com/mediawiki/index.php?title=Facts_%26_Figures

Exponent Inc. (2012). Analysis of Toyota ETCS-i system hardware and software. Retrieved from http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB8QFjAA&url=http%3A%2F%2Fpressroom.toyota.com%2Farticle_download.cfm%3Farticle_id%3D3597&ei=VLMmVY-MCZXtoAT6jYHQBg&usg=AFQjCNEy-ID0BgsyzhBqVoJ0PkCtTbkuXQ&sig2=AIPoXdOf3G660LbMkPZA5A&bvm=bv.90491159,d.cGU

Federal Communications Commission. (2014). April 2014 multistate 911 outage: Cause and impact - report and recommendations. Retrieved from http://www.fcc.gov/document/april-2014-multistate-911-outage-report

Filliâtre, J.-C. (2011). Deductive software verification. *International Journal on Software Tools for Technology Transfer*, *13*(5), 397403. Retrieved from http://dx.doi.org/10.1007/s10009-011-0211-0 doi: 10.1007/s10009-011-0211-0

for Economic Cooperation, O., & (OECD), D. (2013). *Oecd skills outlook 2013: First results from the survey of adult skills*. OECD Publishing.

Games and Learning. (2015). Market analysis. Retrieved from http://www.gamesandlearning.org/category/markets/

Gartner Inc. (2011). Gartner says by 2015, more than 50 percent of organizations that manage innovation processes will gamify those processes. Retrieved from http://www.gartner.com/newsroom/id/1629214

Gartner Inc. (2012). Gartner says by 2014, 80 percent of current gamified applications will fail to meet business objectives primarily due to poor design. Retrieved from http://www.gartner.com/newsroom/id/2251015

Gartner Inc. (2013). Hype cycle for emerging trends. Retrieved 30 April 2015, from http://na1.www.gartner.com/imagesrv/newsroom/images/hype-cycle-pr.png;pv4a3db6f9c029a4db

Gartner Inc. (2014). Hype cycle for emerging trends. Retrieved 30 April 2015, from http://www.pewresearch.org/files/2014/08/FT_gartner-tech-hype-cycle-640px.jpg

Gartner Inc. (2015). *Gartner hype cycle.* Retrieved from http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

Gee, J. P. (2003, October). What video games have to teach us about learning and literacy. *Comput. Entertain.*, *1*(1). Retrieved from http://doi.acm.org/10.1145/950566.950595 doi: 10.1145/950566.950595

Gee, J. P. (2013). Games for learning. *Educational Horizons*, *91*(4), 16–20.

General hype cycle for technology. (n.d.). Retrieved 30 April 2015, from http://en.wikipedia.org/wiki/Hype_cycle#/media/File:Hype-Cycle-General.png

Global education market tops $4 trillion, analysis shows. (2015). Retrieved from http://blogs.edweek.org/edweek/marketplacek12/2013/02/size_of_global_e-learning_market_44_trillion_analysis_says.html

Google trends for the search terms crowdsourcing, citizen science. (2015). Retrieved from http://www.google.com/trends/explore#q=%22citizen%20science

Google trends for the search terms csfv, verigames, formal software verification. (2015). Retrieved from http://www.google.com/trends/explore#q=CSFV%2C%20Verigames%2C%20Formal%20Software%20Verification&cmpt=q&tz=

Gray, J. (1999). Information technology research: Investing in our future.. Retrieved from http://research.microsoft.com/apps/pubs/default.aspx?id=68732

Greengard, S. (2014). DirecTV channels gamification and crowdsourcing. *Baselinemag.com*. Retrieved from http://www.baselinemag.com/innovation/directtv-channels-gamification-and-crowdsourcing.html#sthash.1KxNyC44.dpuf

Grey, F. (2009). Viewpoint: The age of citizen cyberscience. *CERN Courier*. Retrieved from http://cerncourier.com/cws/article/cern/38718

Grier, D. A. (2013). *Crowdsourcing for dummies*. Chichester, West Sussex: John Wiley & Sons.

Guava-libraries: Guava: google core libraries for java 1.6+. (n.d.). Retrieved from https://code.google.com/p/guava-libraries/

Guinard, D. (2011). Epc cloud: Simplifying the internet of things thanks to web patterns: Cloud computing & rest (part 1/3). *Web of Things*. Retrieved from http://webofthings.org/2011/03/08/epc-cloud-1/

Haklay, M. (2013). Gartner's hype cycle and citizen science. *Po Ve Sham - Muki Haklay's personal blog*. Retrieved from https://povesham.wordpress.com/2013/07/08/gartners-hype-cycle-and-citizen-science/

Hanisch, D. E. (2010). *Technology transition and adoption: A study in search of metrics for evaluating transition* (Master's thesis). Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA540127

Hepp, A. (2013, March 12). Dennis crowley @ sxsw 2013 - the future of location (part 1/2) [Video File]. *YouTube*. Retrieved from https://www.youtube.com/watch?v=0rw-SGWieEo

Hern, A. (2014). "internet of things" is the most over-hyped technology, say analysts. *The Guardian*. Retrieved from http://www.theguardian.com/technology/2014/aug/12/internet-of-things-most-over-hyped-technology

Howe, J. (2006). Crowdsourcing: A definition [blog]. *Crowdsourcing, why the power of the crowd is driving the future of business*(December 02).

Imagining the internet. (2015). Retrieved from http://www.elon.edu/e-web/imagining/surveys/2014_survey/default.xhtml

Irwin, A. (2014). From deficit to democracy (re-visited). *Public Understanding of Science*, *23*(1), 71–76.

Jaffar, J., Murali, V., Navas, J. A., & Santosa, A. E. (2012). Tracer: A symbolic execution tool for verification. In *Computer aided verification* (pp. 758–766). Berlin, Germany: Springer.

Kaplan, A. (2010). Intrinsic and extrinsic motivation. *Education.com*. Retrieved from http://www.education.com/reference/article/intrinsic-and-extrinsic-motivation/#E

Kieler, A. (2014). Verizon to pay $3.4 million for not notifying officials of massive 911 service outage. Retrieved from http://consumerist.com/2015/03/18/verizon-to-pay-3-4-million-for-not-notifying-officials-of-massive-911-service-outage/

Kis, V., & Field, S. (2013). Time for the U.S. to reskill? What the survey of adult skills says. Retrieved from http://skills.oecd.org/Survey_of_Adult_Skills_US.pdf

Klein, G., Andronick, J., Elphinstone, K., Heiser, G., Cock, D., Derrin, P., ... Winwood, S. (2010). sel4: formal verification of an operating-system kernel. *Communications of the ACM*, *53*(6), 107–115.

Kling, R., & Leigh Star, S. (1998). Human centered systems in the perspective of organizational and social informatics. Retrieved from http://philfeldman.com/Human_centered_systems_in_the_perspective_of_organizational_and_social_informatics.pdf

Kohn, A. (1999). *Punished by rewards: The trouble with gold stars, incentive plans, a's, praise, and other bribes*. Houghton Mifflin Harcourt.

Kroening, D., & Sharygina, N. (2005). Formal verification of SystemC by automatic hardware/software partitioning. In *Proceedings of the 2nd ACM/IEEE int. conf. on formal methods and models for co-design* (pp. 101–110).

Könighofer, R. (2013). Frama-C: A Quick Start Guide. Retrieved from https://verify.iaik.tugraz.at/teaching/vt/pub/Main/AssignmentThree2013/frama_tutorial.pdf

Lee, J. J., Matamoros, E., Kern, R., Marks, J., de Luna, C., & Jordan-Cooley, W. (2013). Greenify: Fostering sustainable communities via gamification. In *Chi '13 extended abstracts on human factors in computing systems* (pp. 1497–1502). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2468356.2468623 doi: 10.1145/2468356.2468623

Li, G. (2010). *Formal verification of programs and their transformations* (Doctoral dissertation, The University of Utah). Retrieved from http://www.cs.utah.edu/~ligd/publications/Ph.D_Dissertation.pdf

Linden, A., & Fenn, J. (2003). Understanding gartners hype cycles. *Strategic Analysis Report N^o R-20-1971*. Retrieved from http://www.ask-force.org/web/Discourse/Linden-HypeCycle-2003.pdf

Liu, C. (2015). Duolingo: How it can be free to its users forever [Blog]. *HBS OpenForum*. Retrieved from https://openforum.hbs.org/challenge/understand-digital-transformation-of-business/business-model/duolingo-how-it-can-be-free-to-its-users-forever

Logas, H., Whitehead, J., Mateas, M., Vallejos, R., Scott, L., Shapiro, D., ... Lewis, C. (2014). Software verification games: Designing xylem, the code of plants.

M2 Research. (2015). *Gamification 2012*. Retrieved from http://wandameloni.snappages.com/gamification-2012.htm

Marczewski, A. (2012). *Gamification: a simple introduction*. Retrieved from https://books.google.com/books?hl=en&lr=&id=IOu9kPjlndYC&oi=fnd&pg=PA3&dq=Gamification:+a+simple+introduction&ots=kHLp-NlMXZ&sig=qibGMF7bAkcoADcx-CbUEyhzFhs#v=onepage&q=Gamification%3A%20a%20simple%20introduction&f=false

Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., & Crowston, K. (2012). The future of citizen science: Emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, *10*(6), 298–304.

Nielsen. (2012). American families see tablets as playmate, teacher and babysitter. Retrieved from http://www.nielsen.com/us/en/insights/news/2012/american-families-see-tablets-as-playmate-teacher-and-babysitter.html

Nov, O., Arazy, O., & Anderson, D. (2014). Scientists@ home: What drives the quantity and quality of online citizen science participation? *PloS one*, *9*(4), e90375.

O'Leary, D. E. (2008, December). Gartner's Hype Cycle and information system research issues. *International Journal of Accounting Information Systems*, *9*(4), 240–252. Retrieved from https://www.researchgate.net/profile/Daniel_OLeary2/publication/228210690_Gartners_Hype_Cycle_and_Information_System_Research_Issues/links/0deec5169b9a6eace4000000.pdf

Olson, P. (2014). Crowdsourcing capitalists: How Duolingos founders offered free education to millions. *Forbes*. Retrieved from http://www.forbes.com/sites/parmyolson/2014/01/22/crowdsourcing-capitalists-how-duolingos-founders-offered-free-education-to-millions/

Oremus, W. (2014). The never-ending hype cycle - How futuristic technologies go from hyped to hated and back again. Retrieved from http://www.slate.com/articles/technology/history_of_innovation/2014/06/how_the_hype_cycle_explains_moocs_big_data_vr_and_google_glass.html

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. Hoboken, New Jersey: Wiley.

Padua, D. A. (Ed.). (2011). *Encyclopedia of parallel computing*. New York, NY: Springer.

Paharia, R. (2015). 5 gamification trends to watch in 2015. *Bunchball*. Retrieved from http://www.bunchball.com/blog/post/1616/5-gamification-trends-watch-2015

Paliwoda, S., & Thomas, M. (1998). *International marketing*. Butterworth-Heinemann.

Panetta Institute. (2015). Monterey county reads | the panetta institute for public policy. Retrieved from http://www.panettainstitute.org/programs/monterey-county-reads/

Pogue, D. (2015). Apples ResearchKit takes medical research years into the future. *Yahoo.com*. Retrieved from https://www.yahoo.com/tech/apples-researchkit-takes-medical-research-years-117781601479.html

Pope, H., Boaler, J., & Mangram, C. (2015). *Wuzzit trouble: The influence of a digital math game on student number sense* [Pre-publication Draft]. Retrieved from http://www.brainquake.com/wp-content/uploads/2014/04/Pope_Boaler_Mangram_BODY-2.pdf

Poser, S. (2015). Trends in the workplace: Gamification in the enterprise. *Oracle.com*. Retrieved from http://www.oracle.com/us/corporate/profit/big-ideas/012115-sposer-2408614.html

PRWeb. (2012). Global e-learning market to reach US$107 billion by 2015, according to new report by global industry analysts inc. Retrieved from http://www.prweb.com/releases/distance_learning/e_learning/prweb9198652.htm

Radatz, J., Geraci, A., & Katki, F. (n.d.). IEEE standard glossary of software engineering terminology. *IEEE Std*.

Ray, S. (2005). *Using theorem proving and algorithmic decision procedures for large-scale system verification* (Doctoral dissertation, University of Texas). Retrieved from http://www.cs.utexas.edu/~sandip/publications/dissertation/dissertation.pdf

Rimer, S. (2008). Math skills suffer in U.S., study finds. *Nytimes.com*. Retrieved from http://www.nytimes.com/2008/10/10/education/10math.html?pagewanted=all

Rogers, E. (1962). *Diffusion of innovations*. Retrieved from https://books.google.com/books?id=zw0-AAAAIAAJ

Rogers, E. (2010). *Diffusion of innovations* (4th ed.). New York, NY: Simon and Schuster.

Root, A. (2014). Luis von Ahn on duolingo's plans for 2014. *www.crowdsourcing.org*. Retrieved from http://www.crowdsourcing.org/editorial/luis-von-ahn-on-duolingos-plans-for-2014/30191

Rustad, M. (2011). State of literacy in Monterey County. Retrieved from http://literacycampaignmc.org/wp-content/uploads/2011/11/Compressed-State-of-Literacy-MC1.pdf

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*, *55*(1), 68.

Sa'ar, Y. (2010). *Algorithmic methods for formal verification* (Doctoral dissertation, Weizmann Institute of Science). Retrieved from http://ysaar.net/data/PhD_final.pdf

Sauermann, H., & Franzoni, C. (2015). Crowd science user contribution patterns and their implications. *Proceedings of the National Academy of Sciences*, *112*(3), 679684. Retrieved from http://www.pnas.org/content/112/3/679.abstract doi: 10.1073/pnas.1408907112

Schellhorn, G., Wehrheim, H., & Derrick, J. (2012). How to prove algorithms linearisable. In *Computer aided verification* (pp. 243–259). doi: 10.1007/978-3-642-31424-7

Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., ... Bonney, R. (2012). Public participation in scientific research: a framework for deliberate design. *Ecology and Society*, *17*(2), 29.

Shute, V. J., Rieber, L., & Van Eck, R. (2011). Games... and... learning. In R. Reiser & J. Dempsey (Eds.), *Trends and issues in instructional design and technology* (Vol. 3, p. 321-332). Upper Saddle River, NJ: Pearson Education Inc.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge, MA: Massachusetts Institute of Technology.

Sparks, K. (2015). Education through gaming - Inspiration from the White House. Retrieved from https://vungle.com/blog/2015/03/10/education-through-gaming-inspiration-from-the-white-house/

Sprinks, J., Houghton, R., Bamford, S., & Morley, J. (2015). Keeping citizen scientists interested: The importance of task workflow design. Retrieved from http://www.researchgate.net/profile/James_Sprinks/publication/

272176730_Keeping_Citizen_Scientists_Interested_The_Importance_of_Task_Workflow_Design/links/550815050cf27e990e08e6ff.pdf

Sreeraman, S. (2015). Apple brings crowdsourcing to medical research and how!! Retrieved from http://biotechin.asia/2015/03/10/apple-brings-crowdsourcing-to-medical-research-and-how/

Straumsheim, C. (2014). Can Duolingo teach foreign students english? *Slate Magazine*. Retrieved from http://www.slate.com/articles/life/inside_higher_ed/2014/07/carnegie_mellon_partners_with_duolingo_a_language_learning_app_to_make_it.html

Strunk, E. A., Aiello, M. A., & Knight, J. C. (2006). A survey of tools for model checking and model-based development. *University of Virginia*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.78.7581&rep=rep1&type=pdf

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor Books.

Takahashi, D. (2013). With a mobile boom, learning games are a $1.5B market headed toward $2.3B by 2017ă(exclusive). *VentureBeat*. Retrieved from http://venturebeat.com/2013/08/16/with-a-mobile-boom-learning-games-are-a-1-5b-market-headed-toward-2-3b-by-2017-exclusive/

Taylor, V. (2013). 58% of U.S. parents admit to using gadgets to babysit their kids: Study. *Nydailynews.com*. Retrieved from http://www.nydailynews.com/life-style/58-parents-gadgets-babysit-kids-study-article-1.1383009

Tellioglu, U. (2014). *Quantifying the effectiveness of crowd-sourced serious games* (Unpublished master's thesis). Monterey, California: Naval Postgraduate School.

Tellioglu, U., Xie, G. G., Rohrer, J. P., & Prince, C. (2014). Whale of a crowd: Quantifying the effectiveness of crowd-sourced serious games. In *Computer games: Ai, animation, mobile, multimedia, educational and serious games (cgames), 2014* (pp. 1–7).

Toerpe, K. (2013). The rise of citizen science. Retrieved from http://www.wfs.org/futurist/2013-issues-futurist/july-august-2013-vol-47-no-4/rise-citizen-science

Toyota loses first acceleration lawsuit, must pay $3 million. (2013). Retrieved from http://www.autonews.com/article/20131024/OEM11/131029935/toyota-loses-first-acceleration-lawsuit-must-pay-$3-million

Tsotsis, A. (2011). Bing Gordon: Every startup CEO should understand gamification. *TechCrunch*. Retrieved from http://techcrunch.com/2011/06/30/bing-gordon-every-startup-ceo-should-understand-gamification/

Usefulness. (n.d.). *Merriam-Webster's* online dictionary (11th ed.). Retrieved from http://www.merriam-webster.com/dictionary/usefulness

Vehns, M. (2014). *The application of gamification in sales* (Master's thesis, Hochschule für Angewandte Wissenschaften München). Retrieved from http://enterprise-gamification.com/attachments/article/239/The%20application%20of%20gamification%20in%20sales.pdf

Verification vs validation. (2011). Retrieved from http://softwaretestingfundamentals.com/verification-vs-validation/

Verigames.com. (2015). Verigames: Free puzzle games featuring logic, math, and science. Retrieved from http://www.verigames.com

Vesselinov, R., & Grego, J. (2012). Duolingo effectiveness study. *City University of New York, USA*. Retrieved from http://static.duolingo.com/s3/DuolingoReport_Final.pdf

Von Ahn, L. (2006). Games with a purpose. *Computer*, *39*(6), 92–94. Retrieved 2015-02-10, from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1642623

Wasko, M. M., & Faraj, S. (2005). Why should i share? examining social capital and knowledge contribution in electronic networks of practice. *MIS quarterly*, 35–57.

WeWantToKnow. (2015). About us. Retrieved from http://wewanttoknow.com/us/

Wiggins, A., & Crowston, K. (2011, Jan). From Conservation to Crowdsourcing: A Typology of Citizen Science. In *System Sciences (HICSS), 2011 44th Hawaii International Conference* (p. 1-10). doi: 10.1109/HICSS.2011.207

Wilson, F. (2014). Is duolingo a model for freemium education apps? *Gamesandlearning.org*. Retrieved from http://www.gamesandlearning.org/2014/07/29/is-duolingo-a-model-for-freemium-education-apps/

Workman, B. (2013). Gamification: Companies Of All Sizes Are Using This Strategy To Win Customers And Pummel Competitors. *Business Insider*. Retrieved from http://www.businessinsider.com/the-growing-gamification-market-2013-11

Yuen, M.-C., King, I., & Leung, K.-S. (2011). A survey of crowdsourcing systems. In *Privacy, security, risk and trust, 2011 IEEE third international conference on social computing* (pp. 766–773).

ZeroDesktop Inc. (2014). Kids spend more than 3 hours a day on apps [Blog]. *DinnerTime Plus Blog*. Retrieved from http://www.dinnertimeapp.com/blog/2014/09/kids-spend-more-than-3-hours-a-day-on-apps/

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps* (1st ed.). Sebastopol, CA: OReilly Media.

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California